Profiling bacteriophages of the human gut and exploring their application in food and medicine

A Thesis Presented to the National University of Ireland for the Degree of Doctor of Philosophy by

Rhea Lewis, B.Sc.

University College Cork

School of Microbiology

Supervisors: Prof. Colin Hill and Prof. R. Paul Ross

Head of School: Prof. Paul O Toole

February 2020

Table of contents

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Signed: _____ Date: _____

Rhea Lewis

Thesis abstract

Bacteriophages (phages) are viruses which infect bacteria and are ubiquitous in nature. Phage research has undergone periods of interest and indifference, but are now common tools with some suggesting that they will become important as antimicrobial agents in food and in medicine. The first chapter of this thesis is in two parts. Chapter 1a reviews the discovery, basic biology, and applications of phages. They have varied uses such as in models for predator and prey dynamics studies, as a last resort in the fight against antibiotic resistant infections, in the treatment of illness and disease by faecal transplant, in food safety, in phage display and the expression of antibodies and peptides, and in metagenomics studies to understand human health and disease. In the chapters that follow four of these applications were investigated. Chapter 1b addresses the barriers that should be taken into account in order to use phages successfully in food and feed. These issues can also arise when using phages in phage therapy.

Chapter 2 investigated a commercially available phage against *Listeria monocytogenes*, the causative agent of a serious foodborne illness. Phages can be used as clean label additives which are becoming more accepted than traditional additives by consumers. Phage alone and in combination with a commercially available bacteriocin, a ribosomally produced antimicrobial peptide, caused a significant reduction of *L. monocytogenes* in a food model.

In Chapter 3 the effect of hormonal contraception on the human faecal virome was investigated in a pilot study. No differences in viral or bacterial diversity were found between men and women but viral diversity was found to be reduced in women using hormonal contraception compared to women not using hormonal contraception.

Clustering increased the discriminatory power of the virome data. This could be important for establishing inclusion or exclusion criteria for subjects or choosing healthy controls for studies.

Chapter 4 addresses the isolation of phages and the analysis of single nucleotide polymorphisms (SNPs) in phage strains. Notwithstanding the importance of culture-independent methods for studying phages the isolation of phages is still necessary. No phages were isolated from the same source as bacterial strains but a number were isolated from environmental samples, illustrating the value of looking in different sources and new environments. 70 strains of *Escherichia coli* phage APCEc01 were isolated and compared to each other and the original sequenced genome of APCEc01. This highlighted the variation that occurs in genomes of the same phage and its possible impact.

In Chapter 5 a jumbo phage was isolated against *Klebsiella aerogenes*, a member of the ESKAPE pathogens which are of concern due to their ability to cause hospital acquired and antibiotic resistant infections. The phage isolated was effective against *K. aerogenes* biofilms which are routinely found on medical devices. The phage in combination with Phage K was effective against a mixed biofilm of *K. aerogenes* and *Staphylococcus aureus*.

Phages have myriad applications in food and medicine. Since their discovery the number and breadth of applications have continued to grow. This research highlighted these uses and investigated a number of them in greater detail. This included the use of phage as useful tools in food safety, the first study highlighting the effect of hormonal contraception on the human gut virome, the isolation of a novel phage against a clinically relevant pathogen and its application as a tool to combat

biofilms, and an investigation into the isolation of phages and the incidence of single nucleotide polymorphisms in strains of phages.

Chapter 1a

The use of bacteriophages in food and medicine

Introduction

Bacteriophages (phages) are viruses of bacteria and therefore require a sensitive bacterial host for the production of new phages. Phages are the most abundant biological entities on earth and are found in almost every environment including the guts and homes of humans, in the guts of insects, and in oceans (1-4). They are also found in more extreme environments such as in hot springs, deep sea thermal vents, and arctic sea ice (5-7). Some estimates put viral numbers at 10 times that of bacterial cells worldwide (8). In some environments viral numbers can be up to 150 times greater than bacterial numbers. Viral abundances in oceans vary depending on the depth and environment, with the majority between $3.7X10^5$ and $6.4X10^7$ per ml. Virus-like particles were estimated to be present in hot springs at concentrations ranging from $7X10^4$ to $7X10^6$ per ml (5). Viral abundance in human faeces has been estimated at $1.1X10^8$ to $1.8X10^{10}$ per gram (9, 10).

Discovery of phages

Phages were independently discovered by Frederick Twort and Felix D'Herelle in 1915 and 1917, respectively (11). Twort discovered phages while carrying out experiments to investigate how viruses could be grown without the use of host cells. Twort believed that viruses from sources such as water, vegetation and faeces after incubation in water for periods of time and subsequent filtration could be grown outside of cells on artificial media consisting of an agar, egg or serum base with the addition of chemicals or extracts. Although this is now known to be false this experimentation led to the discovery of phages. After inoculating agar with unfiltered fluid used for vaccinating against small pox micrococcus contamination grew while

vaccinia virus, unsurprisingly, did not. It was observed that some micrococcus colonies took on a “glassy and transparent” appearance and could not be sub-cultured and grown on any media (12). These “glassy and transparent” colonies passed the attributes on to any colonies in which they came in contact and could be filtered and still retain this ability. Twort had discovered a virus of bacteria but was not confident in naming it so and simply said that it might be “an acute infectious disease of micrococci". The work was interrupted by Twort’s service during World War 1 (13). He did not continue with the investigation of what we now know to be bacteriophages after he returned to research.

D’Herelle first viewed what we know to be phage plaques when culturing coccobacillus to infect and kill plagues of locusts as areas free of growth in confluent lawns of bacterial cells (14). He came to the opinion that two agents were present, one infecting the locusts and one inhibiting the growth of the bacterial culture that was filterable. The true discovery of phages by D’Herelle occurred when he was investigating a dysentery epidemic (11). Encouraged by his coccobacilli work D’Herelle followed the appearance of plaques from the filtered faeces of dysentery patients on cultures of bacteria isolated from the faeces of dysentery patients. At first he considered the clear spots could be the causative agent of the dysentery but reconsidered this after the feeding of filtered faeces of infected individuals failed to cause an infection in animals. He observed that plaques occurred during the recovery of patients and that the filterable agent could make a turbid bacterial culture clear. He had discovered a filterable virus that was parasitic on bacteria and named them bacteriophages. D’Herelle investigated them for their use in phage therapy and developed techniques which are still in use today. The areas free of growth seen in the coccobacillus cultures were not mentioned in his first communication about phages

but instead concentrated on his work regarding dysentery. D'Herelle denied having prior knowledge of Twort's work and asserted that he only became aware of it in 1921 after it was brought to his attention by other scientists.

Phage biology

Structure and taxonomy of phages

Phages share a common structure of a filamentous nucleic acid core surrounded by a protein coat (15). The nucleic acid core can differ in being double-stranded or single-stranded and consist of DNA or RNA. The protein coat consists of repeated identical subunits called capsomeres which give the virion a symmetrical structure which is prolate or icosahedral in shape. Icosahedral heads are essentially round in shape and are made up of identical capsomeres while prolate heads resemble an elongated icosahedral shape. The identical, repeated capsomeres reduce the number of different proteins which must be expressed by the phages to form new functional virions which is necessary as phages often have genomes which are smaller than would be expected to be necessary for the production of progeny. The entire infective particle is known as a virion. In general, phages consist of a head attached via a neck, surrounded by a collar, to the tail sheath and base plate which can then end in tail spikes or fibres (16, 17). Some do not have tail structures such as the *Microviridae*, which have circular head structures with no tail, or filamentous phages which are long and rod-shaped (15).

Although the 16S rRNA gene can be used to classify bacteria phylogenetically phages have no conserved genes which can be used in a similar way (18). Phages have been historically classified based on a number of different parameters such as nucleic

acid composition, phage morphology and the bacterial host they infect. Each method of phage classification has its disadvantages, such as the lack of homology if using conserved genes and a lack of distinction using nucleic acid and morphological based classifications. The current taxonomy system for phages groups them together based on shared characteristics with further sub-division into families based on more closely related shared characteristics. Genus and sub-genus are characterised by genome configuration, host range and genome size. This system fails to take into account new information from metagenomics and proteomics and has led to calls for changes to reflect the growing field (18, 19). For example, despite the fact that they are so closely genetically related that they form functional hybrids P22 and lambda phage are placed in different families. It has been suggested that metagenomic sequencing data should be included in the taxonomy criteria.

The order *Caudovirales* are double-stranded DNA tailed phages which consist of the families *Myoviridae*, *Siphoviridae* and *Podoviridae* among other phage families (Fig. 1) (20). *Myoviridae* have long contractile tails, *Siphoviridae* have long noncontractile tails and *Podoviridae* have short noncontractile tails. Phage tails are involved in the recognition of host cells, adsorption to the bacterial cell, penetration of the cell, and delivery of nucleic acid from the head into the bacterial host cell (21). Phage tails can end in a tail tip or more complex base plate structure with tail fibres or spikes which initiate infection after the tail fibres or spikes bind to the host cell. Phage receptor-binding proteins recognise bacterial protein or carbohydrate receptors on the bacterial cell. Common receptors on Gram negative bacteria include LPS, flagella proteins, pili, porins and other outer membrane proteins. Peptidoglycan and teichoic acid, components of the cell wall, are common receptors in Gram positive bacteria (22). Bacterial resistance to phage can occur by bacteria modifying their receptors but

also by the degradation of invading phage nucleic DNA (21). In turn phage can modify their receptor-binding proteins to overcome bacterial resistance. Phage receptor-binding proteins are highly specific in their recognition of bacterial receptors and in this way play an integral role in determining the host range of a phage in conjunction with factors such as the presence of protective plasmids or prophage, phage resistance mechanisms and biochemical interactions (23). The host range of a phage can vary from a single bacterial strain to multiple bacteria from different genera. Phages are often described as having a broad or narrow host range but these terms are ambiguous in their description of the host range. Two phages infecting a number of strains or a number of species could both be described as having a broad host range. After recognition of receptors signalling occurs through the tail fibres or tail tip to the tail causing the ejection of nucleic acid from the head through the tail and into the host cell (24). In the case of T4 phage, a member of the *Myoviridae* family which all contain a sheath around the tail, the long-tail fibres recognise *Escherichia coli* and short-tail fibres located on the base plate bind irreversibly to the host cell causing a conformational change in the base plate from a high energy conformation to a low energy conformation. This energy change contracts the tail sheath causing the tail tube to enter the host cell where the phage DNA is introduced.

Single-stranded DNA phages are grouped into eight families with two families infecting bacteria, *Microviridae* and *Inoviridae* (25, 26). Both families have circular genomes and replicate by rolling-circle replication. *Microviridae* have small icosahedral capsids while *Inoviridae* are filamentous phages. The *Microviridae* are further sub-divided into the genus *Microvirus* and the sub-family *Gokushovirinae*. Single-stranded DNA phages are less widely studied than the *Caudovirales*. The International Committee on Taxonomy of Viruses (ICTV) phage database contains

predominantly *Caudovirales* phages and also more *Inoviridae* phages than *Microviridae* phages. Conversely MetaVir, a web server for annotation of viral metagenomics datasets, contains a much higher percentage of single-stranded DNA phages than the ICTV database with the majority of these being *Microviridae* phages rather than Inoviridae phages.

RNA phages are grouped into two families, *Leviviridae* and *Cystoviridae* (27). The *Leviviridae* are single-stranded RNA phages while the *Cystoviridae* are segmented, double-stranded RNA phages. Very few genomes are available for either RNA phage family, in part due to the focus on DNA phages in metagenomics but also because of difficulties in culturing RNA phages. Difficulties in culturing also occur with DNA phages. RNases in human and animal samples can degrade or damage RNA phages and make their discovery or study less likely (28). For this reason it has been suggested to separate RNA and DNA from samples and to eliminate RNases. There are also suggestions that some viruses have been misidentified as RNA viruses when they may actually be RNA phages. This draws attention to the need for better characterisation of RNA phages. Single-stranded RNA genomes can be negative or positive sense depending on their orientation and whether they require transcription before translation. Positive sense RNA can act directly as mRNA during replication.

Phage genomes are often small and only consist of a few genes necessary for the replication of the phage genome and protein coats. Phages use the machinery of their host for replication. RNA phage genomes can reach up to ~4kb in length which is considerably smaller than double stranded DNA phages of which the smallest is ~11kb (27, 29). Tailed phages with genomes larger than 200 kbp are termed jumbo phages (30). Jumbo phages are characterised by their large genomes and large virions. Tailed phages require at a minimum genes for DNA replication, transcription

regulation, DNA packaging, head, tail and tail fibres, and lysis (31). In addition to these core genes phages can carry genes to improve or aid phage and bacterial host survival. Bacterial virulence factors and toxins can also be phage encoded.

Phage replication

Phages require the machinery of a bacterial cell to express the proteins required to assemble new phages. The phage life cycle involves adsorption to a host cell, injection of nucleic acid into the host, replication of nucleic acid and production of phage proteins using the host machinery, virion assembly, and release (Fig. 2) (32). Phage life cycles vary and are described as lytic, lysogenic, pseudolysogenic, or chronic. Phages can be described as virulent or temperate with virulent phages undergoing a progeny phage producing lytic lifecycle while temperate phages can change between a progeny phage producing lytic lifecycle and a non-producing lysogenic lifecycle (33). As the name implies lytic replication requires the lysis of the bacterial host cell for the release of progeny phages (34). A lysogenic life cycle occurs when the phage genome is integrated as a prophage into the bacterial genome (33). Phage proteins are then expressed during the normal expression of the bacterial genes along with the bacterial proteins. Prophages have been found in ~60% of published bacterial genomes. A lysogenic life cycle can also occur when phages survive as plasmids in the host cell (34). Lysogeny may be useful when too few encounters will happen between phages and a permissive host, for instance when host abundance is too low or when too many phages are being removed from the system (32). This can occur when resources are limited (33). Phages use repressor genes to control the switch from lysogeny to lysis by repressing the lytic cycle. These repressors can also block the integration of any other phage genomes. The infection of a bacterial cell by multiple phages can occur and is called superinfection. The induction of prophages

can be triggered by impending host death which can be seen as DNA damage. Bacterial DNA damage can be caused by a range of chemical and physical methods including antibiotics targeting DNA replication, UV damage and reactive oxygen species. Prophage induction can also be triggered by molecules which signal that the phage is in a better location to find a receptive host (35). Cell lysis takes place using endolysins to break down the peptidoglycan and holins to damage the plasma membrane (32). When no cell lysis occurs phages can be released by budding or extrusion in what are termed chronic infections where phages are continually released. Pseudolysogeny is a poorly understood system whereby phage nucleic acid occupies the cell in a non-active state and does not establish a lysogenic or lytic lifecycle (36). This may be due to a starved or stressed bacterial state which does not provide enough energy or the correct conditions for phage replication.

Phages in any form, lytic or lysogenic or something in between, can aid their host (37). Free phages kill related competitors allowing their hosts access to resources. Infection by prophages can protect hosts from infection by closely related phages. Phages are important for horizontal gene transfer between bacterial species and act as reservoirs of genetic diversity (38). Host bacterial genes can be accidentally packaged in phage heads instead of phage DNA. These phages lacking phage genes can then attach to new hosts and deliver the bacterial DNA that is then incorporated into the new genome. Also prophages can carry genes that are incorporated into the host genome with the prophage. Botulism, diphtheria, cholera, and illnesses caused by Shiga toxigenic *E. coli* are all caused by exotoxins, all of which are phage-encoded virulence factors. Virulence factors are not required by the phages for replication but aid the bacteria in proliferation and infection of a mammalian host. Phages can also carry antibiotic resistance genes between bacteria (39).

Bacterial resistance to phages

On the other hand, phages can also hinder bacteria as evidenced by the many methods that bacteria use to stop phage infection (40, 41). Bacteria can resist phage infection by a number of methods such as preventing phage adsorption, superinfection exclusion, restriction-modification, CRISPR-cas, and abortive infection. Bacteria can stop or reduce adsorption by phages by modifying or completely removing receptors, reducing the number of receptors, masking receptors by the production of extracellular matrix, or the production of competitive inhibitors. Low phage adsorption rates can cause changes in phage dynamics. In phage dynamics models the rate at which the phage and bacterial populations change depends on five parameters - the infectivity of the phages, which is based on the adsorption rate; the burst size, which is the number of phage progeny produced from a single bacterial cell; the latent period, which is the time between adsorption and release of progeny phages; the rate at which the phages are inactivated or removed from the site of the infection; the maximum bacterial growth rate and two variables - the density of susceptible bacteria; and the density of phages (42).

Superinfection exclusion is counted as a phage resistance system but is due to the presence of a phage in the host genome. It occurs when the presence of a phage in a host cell prevents infection by a second phage (43). Superinfection exclusion is carried out by the phage to avoid lysis of the host cell by the invading phage and the original phage being lost in the process.

Bacteria use restriction-modification systems to protect against invading DNA in two ways using methylase and restriction endonuclease enzymes (41). Host DNA is methylated to protect against restriction. Upon entry into the cell phage DNA will

be recognized by a restriction enzyme, as it is unmethylated, and degraded. In some cases foreign DNA can be methylated upon infection by a bacterial methylase and avoid restriction.

Bacteria can also use the CRISPR-cas system (Clustered Regularly Interspaced Short Palindromic Repeats loci, coupled to CRISPR-associated genes) for phage resistance (44). Upon interaction with a phage the bacteria integrates spacers derived from the phages DNA into the CRISPR locus between CRISPR repeats. Bacteria can insert multiple spacers to derive widespread resistance. The CRISPR locus provides specificity while the cas system provides the enzymatic function and degrades phage DNA. Bacteria use multiple systems together to protect against phage infection such as a combination of restriction-modification and CRISPR-cas (45).

Abortive infection provides resistance by the abortion of phage infection; however, this also kills the infected cell (46). This limits the spread of phage by decreasing the number of progeny released. Abortive infection systems can interfere with phage DNA replication, RNA transcription, phage development, and morphogenesis (47).

Phage overcoming resistant bacteria

Phage can also try to overcome bacterial phage resistance (48). Firstly, phage resistance itself can reduce bacterial fitness and the ability of bacteria to compete with phage-sensitive bacteria. Phage resistance can reduce the virulence of bacteria due to the loss or mutation of virulence factors (49). Through mutation phages can modify their receptor binding proteins to target alternative proteins to combat changing receptors (50). Phages can produce enzymes to break down extracellular matrices that protect receptors from recognition. Phages can modify restriction sites or place them

in an unrecognisable order, stimulate methylase enzymes to cause methylation of phage DNA so it is recognised as bacterial DNA, or modify DNA to avoid recognition by restriction modification systems. To avoid the CRISPR-cas system phages can use mutations in the protospacer or protospacer-adjacent motif or produce anti-CRISPR proteins that interfere with the system. Toxin-antitoxin systems are a subgroup of abortive infection systems that contain a toxin that induces cell death or dormancy and a neutralising antitoxin. Phages can avoid these systems by neutralising the toxin or mimicking the antitoxin. In other cases phages can prevent the activation of abortive infection systems or exchange genomic regions with prophages that are already present.

Applications of phages

Phage dynamics

Phages can change the composition of bacterial communities through the predation of bacteria. Bacterial and phage co-adaptation leads to evolution of community members and aids in the maintenance of a diverse community and a large pangenome (51, 52). In the absence of phages bacterial diversity is dependent on the availability of substrates and bacterial substrate usage. Changes in substrates can cause diversity to be low as less adapted bacteria will be out-competed and users of the most commonly available substrate will overgrow. The presence of phages allows for the exploitation of less favoured niches and multiple bacterial lineages will be able to compete causing higher bacterial diversity.

Antagonistic coevolution is the reciprocal development of predator and prey resistance and infectivity in response to one another (53). Antagonistic coevolution

increases community diversity. Phages and bacteria are often used as models for the study of antagonistic coevolution. Simplified benchtop experiments can be used to try and examine the importance and effects of phage dynamics in the gut. Phages increase bacterial fitness, by superinfection and introducing virulence or resistance genes, and competition. Phages can increase rates of bacterial evolution through stress response and genetic mutation (54). This increased evolution caused by the presence of phages has been suggested as a reason for the inter-individual variation seen between individuals (52, 55). Lysogenic cells have been found to be at a disadvantage in unstructured habitats and at an advantage in structured habitats, which could represent different areas of the human gut (56). Inaccessible areas such as crypts, biofilms, and the appendix could act as refuges for bacteria that prevent complete elimination and because of this prophage may have more of an effect than lytic phages (57).

Phage therapy

Although there was a period of interest in using phages to combat infectious disease, such as in the control of dysentery and during World War II, this ended due to the discovery and use of antibiotics (58). This was also due to the lack of strong scientific experimentation and literature for the use of phages as phage therapy and the lack of basic knowledge about phages (59). The majority of evidence for the use of phage therapy was small scale and based on largely anecdotal accounts. Another important issue was the efficacy of antibiotics, in that while they were so effective in the treatment of illness and saved the lives of so many it may have seemed unnecessary to look for an alternative (60). Political issues also had an effect on the use and trust placed in phage therapy. Phage therapy was mostly used by the Soviet Union and the German army during World War II and this lead to distrust around the use of it by the western world. Although phages have been studied deeply the lack of knowledge about

phages is still an issue when it comes to using phage therapy. Phage therapy is still in use in Georgia, Poland and Russia.

Some estimates suggest that by 2050 ten million people will die every year due to antimicrobial resistant infections (61). The WHO has recently declared antibiotic resistance a global emergency, with alternatives treatments required immediately (62). In a 2015 Global Action Plan on antimicrobial resistance the WHO described antimicrobial resistance, which includes antibiotic resistance, as a threat to "the very core of modern medicine and the sustainability of an effective, global public health response to the enduring threat from infectious diseases". The action plan also highlighted the shortage of emerging antimicrobials which could bring us from a pre-antibiotic era to an antibiotic era to a post-antibiotic era in a matter of 100 years. The WHO pledged US$41.7 million to combating antimicrobial resistance for one financial year alone. The WHO has compiled a list of pathogens that are a priority in the fight against antimicrobial resistance (63). *Mycobacterium tuberculosis* is the top priority pathogen with carbapenem-resistant *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* described as critical priorities. Vancomycin-resistant *Enterococcus faecium*, clarithromycin-resistant *Helicobacter pylori*, fluoroquinolone-resistant *Salmonella* species, vancomycin-resistant or methicillin-resistant *Staphylococcus aureus*, fluoroquinolone-resistant *Campylobacter* species and 3rd generation cephalosporin-resistant and fluoroquinolone-resistant *Neisseria gonorrhoeae* are described as high priority pathogens. Penicillin-non-susceptible *Streptococcus pneumoniae*, ampicillin-resistant *Haemophilus influenzae* and fluoroquinolone-resistant *Shigella* species are described as medium priority pathogens. The low number of new antibiotic compounds in development and the long time frame associated with their development and approval are of concern. The low

likelihood of approval means that only one or two in ten antibiotics will be approved for market and the average development time from Phase-1 safety trials to approval is about seven years. Many antibiotics in development are derivatives of well-established antibiotics and resistance will develop quickly. The low success rate of approval and the paucity of new classes of antibiotics means these will not be enough to combat the global antimicrobial resistance issue. The WHO report on antimicrobial resistance also draws attention to the importance of antibiotic use in animals adding to antimicrobial resistance. Antimicrobial resistance is a One Health problem meaning that the health of humans, animals and the environment are connected (64). Environmental bacteria can act as reservoirs of antimicrobial resistance genes that can be transferred to humans or animals. These genes can enter the environment through human and animals waste but also through the pharmaceutical industry. Combating a One Health problem requires an integrative approach through targeting of all of these areas.

Research is growing in the use of phages as medical interventions. Phage therapy is a potential solution to antibiotic resistant bacteria for which all antibiotic options have been exhausted (65, 66). Historically the fact that phages do not affect a broad spectrum of bacteria like antibiotics was seen as a disadvantage of their use, but is now seen as a positive attribute due to their more limited impact on the surrounding microbiota and in fighting against the rise of antibiotic resistance. They are self-replicating meaning small numbers of phages can be added and numbers will increase if the pathogen is present. Conversely the self-limiting nature means that if host numbers decrease phage numbers will also decrease. The disadvantages associated with phages can be overcome by selecting the correct phages. Phages should be lytic and non-temperate, and therefore unable to transfer virulence or antibiotic resistance genes. Phages should be nontoxic to the organism they are being used in and no toxins

or toxin genes should be present. To increase the host range of the therapy and to reduce the problem of resistance phages can be used in cocktails. Phages could also be used in combination with antibiotics (67). Although not many phages have undergone clinical trials, phages against a range of the priority pathogens have been used in phage therapy (68-71) and in-vivo models (72-76).

Phages have been effective in compassionate use cases and individual cases in clinical trials but overall clinical trials did not prove efficacy of phages (77). For phages to be used commercially they will need to be produced under Good Manufacturing Practice (GMP) guidelines and infrastructure. This could be problematic as GMP production would greatly increase the cost. In the case of PhagoBurn, a Phase I and II clinical trial carried out in using a phage cocktail against burn wound infections, GMP production was suggested to be the cause of delayed patient recruitment and reduced phage titre.

Endolysins

Endolysins are enzymes that degrade peptidoglycan, a constituent of the bacterial cell wall, produced late in the lytic phage life cycle of double stranded DNA phages to allow for the release of progeny phages (78). Peptidoglycan is composed of repeat polymers of the amino sugars N-acetylglucosamine and N-acetylmuramic acid, linked by β-1, 4 glycosidic bonds. Tetrapeptide side chains are attached to the lactyl group of the muramic acid by amide bonds. Endolysins are hydrolase enzymes and can be separated in to five groups N-acetylmuramidases (Iysozymes), endo-β-N-acetylglucosaminidases, and lytic transglycosylases, which all cleave the glycosidic linkage of peptidoglycan; endopeptidases, which cleave the peptide moiety; and N-

acetylmuramoyl-L-alanine amidases, which cut the amide bond between both moieties.

Endolysins have been suggested as tools as opposed to the use of whole phages. Endolysins have the advantage that they are modular in nature making engineering possible (79). Endolysins can be engineered for enhanced ability or host range. No resistance to endolysins has been detected (80). Endolysins also have the advantage of having increased host range compared to corresponding whole phages and temperate phages that are not suitable for phage therapy can be used as endolysins (81). Also the use of purified phage endolysins could avoid any issues of immune response caused by the use of whole phages. Endolysins have been effective in the treatment of natural cases of bovine mastitis caused by *S. aureus*, MRSA is a WHO high priority pathogen. Endolysins were also effective in the treatment of experimentally induced *Streptococcus pneumoniae*, a WHO medium priority pathogen, pneumonia in mice (82). Endolysins have been suggested as a treatment for *Clostridium difficile* infection (CDI), a common consequence of antibiotic therapy that is currently treated with antibiotics and commonly recurs (83). No lytic *C. difficile* phages have been isolated to date meaning they are unsuitable for phage therapy. Endolysins have been investigated for use in food safety, in foods such as soya milk and lettuce, similarly to whole phages to reduce contamination with bacterial pathogens (84). Endolysins have also been investigated for use in detecting bacterial pathogens, especially in food. Current methods for detection include plating on selective media, PCR, and antibody based detection. These methods can involve long incubation periods and expensive molecular techniques. The detection limit for antibody based detection methods can be high with low specificity of binding for pathogens. Endolysins contain conserved cell wall binding domains which recognise

their specific host. Cell wall binding domains can be fused to fluorescent proteins and used to probe for pathogens. They act as rapid and sensitive detection systems (80).

Faecal transplant

Faecal transplants, the transfer of whole faeces or portion of the microbiota contained within it, are gaining interest for the treatment of human illnesses and conditions (85). Faeces is a complex collection of bacteria, plant viruses, animal viruses, phages, archaea, fungi, human colonic epithelial cells, protists, and metabolites produced in the human body. Metabolites are often produced by commensal bacteria rather than human cells. It is not clear what component is the active agent of changes associated with faecal transplants. Bacterial numbers have been estimated at nearly 10^{11} bacteria per gram of wet stool, but almost 50% of this number could be injured or dead cells. Viral numbers have been estimated at 10^{8} to 10^{9} viruses per gram of wet stool but could be up to 10^{10} viruses per gram of stool (86). However, phage concentrations are greater in the mucous associated with the gastrointestinal tract than the gut lumen and therefore the faeces (37). Phages may act as non-host derived immunity for the mucosal surfaces by lysing or binding bacteria and reducing microbial colonization.

Faecal microbiota transplants (FMT) have been mostly used to treat CDI. The role of the virome has also been assessed in some of these cases. Draper *et al*. (87) and Zuo *et al*. (88) both found an increased abundance of *Caudovirales*, a decreased abundance of *Microviridae*, and an increased abundance of *Anelloviridae* in patients suffering from CDI compared to healthy controls. CDI is characterised by decreased bacterial diversity due to an overgrowth of *C. difficile*. It is possible that the success of FMT is not due to the presence or absence of particular members of a community

but is due to the introduction of a more diverse community of which phages are a significant part. The failure of FMT may be due to inadequate engraftment of the donor microbiota. Zuo *et al.* (88) found that FMT responders acquired more donor-derived contigs than non-responders. The importance of donor selection was also called to attention as a positive treatment outcome was more likely if the *Caudovirales* richness of the donor was higher than that of the recipient. In a follow up study to the treatment of CDI with FMT it was suggested that FMT could be regarded as much a viral therapy as a bacterial therapy as viruses are such a large component of faeces (89). In fact, sterile faecal filtrate has been used to successfully treat CDI, albeit in a small number of cases (90). The filtrate contained bacterial debris, viruses, proteins, antimicrobial compounds, metabolic products, and oligonucleotides. Patients were followed up for between six months and two years and five months and all remained symptom-free for the entire study period. More study is required in the long term effects of FMT to pinpoint what is happening to the microbiome that allows for remission of disease and what components are causing the changes. In addition to CDI FMT has been studied in the treatment of immune checkpoint inhibitor therapy-associated colitis, ulcerative colitis and in the eradication of multidrug-resistant *Enterobacteriaceae* in the gut (91-93).

The choice of donors is of the utmost importance as faecal transplant can have unwanted or unexpected effects. Recently it was reported that two individuals who had received FMT from the same donor developed invasive infections, caused by extended-spectrum beta-lactamase producing *E. coli*, of which one was fatal (94). This has led to the introduction of new screening requirements for donors by the FDA (95). These new requirements require samples to be tested for multidrug resistant organisms, and subsequent exclusion if they are positive; exclusion of people at a

higher risk of multidrug resistant organisms carriage from being donors; screening of stored donor faeces before use; and for informed consent to include the risks of FMT, the risk of multidrug resistant organisms, and the donor screening and testing process. It would be possible to carry out these checks on donors of transplants composed solely of the viral fraction.

Food safety

Twelve phage preparations have been approved for use in food in the U.S. and granted Generally Regarded as Safe (GRAS) status by the U.S. Food and Drug Administration (Table 1) (96). These can be used to target a selection of pathogens commonly associated with food including *Listeria monocytogenes*, *Salmonella*, *Shigella*, and *E. coli*. Some are licensed for use in all foods, such as ListShield or SalmoPro, while other are licensed for use in specific foods such as PhageGuard which can be used in beef only.

Phage display

Phage display is the insertion of foreign DNA fragments into a gene of the phage coat protein and the subsequent expression of the foreign DNA fragments on the phage surface (97). Phage display can be used for antibodies or peptides. Phages displaying the peptide or antibody of interest are then selected by binding to the antigen of interest and washing to remove non-binding peptides or antibodies. Higher affinity binding peptides or antibodies can be selected by specific wash steps. Phages are then eluted and amplified using *E. coli* and can be used in a subsequent round of selection or analysis. An advantage of phage display is the direct linking of phenotype of the displayed peptide to the genotype. Phage display also lends itself to high throughput screening, through the creation of random libraries and screening, which

has led to its use in developing diagnostics and therapies. Phage display can be used to screen peptide libraries and whole cells (98). The screening of whole cells has many advantages such as cellular receptors are kept in their native state, peptides that mediate cellular uptake can be preferentially selected by washing, there is no selective pressure towards a specific macromolecule, and no previous knowledge is needed about which macromolecules to target. A common target for whole cell phage display screening are cancer specific peptides on tumours. Once isolated the tumour specific peptides can be used for tumour specific drug delivery by the attachment of the peptide directly to chemotherapeutic agents or indirectly to a drug carrier. They can also be used for in vivo and in vitro tumour diagnostics.

Phage display can be used to identify bacterial surface proteins by screening for mimotopes (99). Mimotopes are mimics of epitopes and cause similar antibody responses to their epitope. This can be used for vaccine development when little is known about the pathogenesis of a bacteria or its surface proteins have not been identified. Mimotopes can be used to identify the epitope they are similar to by homology search to the bacterial genome.

Phage display technology can be used to produce monoclonal antibodies and has many advantages over traditional hybridoma technology (100). Phage display does not require the use of animals, can be used to produce antibodies against toxins and conserved antigens which hybridoma technology cannot, and is also faster than hybridoma technology. Phage displayed antibodies can be easily engineered and will not contain any mouse antibodies which can cause reactions when used in humans. Adalimumab (Humira®) was the first antibody generated using phage display approved for use; it became the bestselling antibody drug and achieved more than ten billion USD global sales in 2013 (101). It is used in the treatment of inflammatory

autoimmune diseases by the binding of tumour necrosis factor to inhibit the proinflammatory cytokine cascade.

Viral metagenomics studies

As previously mentioned, there is no conserved gene that can be targeted for phage identification like the bacterial 16S rRNA gene. Although advances in molecular biology and metagenomics have increased knowledge of phages there is still a lot to be learned and some issues remain to be resolved (25). Shotgun metagenomics can be useful in cases where phages cannot be cultured, which often occurs when the host itself cannot be cultured. In metagenomics studies there can be a bias towards certain groups of phages based on the methods used. Even within the relatively small group of eight families of single stranded DNA phages considerably more is known about one family, the *Microviridae*. Very few RNA phages have been isolated and sequenced and they are rarely found in metagenomic studies. This bias can be due to a lack of RNA and ssDNA reference genomes that allow us to assign taxonomy to sequences but also some virome extraction and library preparation methods can bias the system. In one study a linker amplified shotgun library returned only double-stranded DNA viruses while multiple displacement amplification returned mostly single-stranded DNA viruses using the same sample (102). Viral DNA extraction methods can also affect sequencing outcomes (103). There are large numbers of sequences in metagenomic studies that cannot be identified or taxonomically categorised, this can be referred to as "viral dark matter" and can represent 60-95% of sequences in datasets (104). This may be due to the absence of a universal gene marker for phages and the relatively low number of phage sequences in publically available databases, in comparison to bacterial sequences. These unidentified sequences can make it difficult to mine information from data leading to

issues such as not being able to link bacterial hosts to phages or understanding the impact that phages have on a community. Another problem associated with viral metagenomics is the inability of some assembly programs to accurately assemble virome data (105). In some cases viromes simply cannot be assembled while in other cases multiple viromes can be misassembled together to form chimeric sequences that will be unidentifiable. Difficulties in assembly can be caused by bias in the processing and sequencing of viromes, unequal read coverage, the highly variable genome structure of phages due to horizontal gene transfer, and contamination by and subsequent inclusion of bacterial or human reads. The inclusion of prophage sequences in bacterial genomes and bacterial DNA in phage genomes from previous hosts can further complicate annotation (38, 39). Viral load has been shown to vary greatly between individual samples (86). Few studies have taken this in to account and often just rely on relative numbers rather than absolute. The correct choice of virome preparation protocol, assembly program and database are essential to an accurate viral metagenomics study.

Metagenomics can be used to find things which were previously unknown such as crAssphage which was discovered using metagenomics data in 2014, despite never being cultured (106). CrAssphage has been found to account for up to 90% of the reads in the virus-like particle-enriched fraction of some gut metagenomes studies (107). Subsequently, crAssphage has been suggested to be a family of phages rather than a single phage as previously indicated. In 2018 a crAssphage was cultured for the first time and its host was confirmed as *Bacteroides intestinalis* (108).

The human virome is highly specific to individuals (109). This makes looking for trends and changes difficult as the viromes of subjects vary so much. While the bacteriomes of monozygotic twins and their mothers were more similar than those of

unrelated individuals the viromes did not reflect this and were unique to individuals. While inter-individual variation was high, intra-individual variation was low over the period studied. This individuality was also seen by S. Minot et al. (55) who showed that over a two and a half year period 80% of viral contigs persisted in an adult subject. Phages were defined as "core" if found in more than 50% of people in a dataset, "common" if found in 20–25%, and "rare" representing phages rarely shared or unique to a person (110). Together the "core" and "common" phages were proposed as a healthy gut phageome. The selection of healthy controls in studying disease is important and it has been suggested that household controls would be the best choice as they would be the most similar (111). This was also highlighted for microbiome studies where members of a household were more similar than those not sharing a home (2). Viral metagenomics studies could be useful in identifying the cause of illnesses or as biomarkers of disease. For example, a significant increase in viral richness was seen in IBD sufferers compared to healthy controls (111). The viruses causing this increased richness differed for ulcerative colitis and Crohn's disease sufferers. V. Pérez-Brocal et al. (112) found an association between some viruses and the onset of inflammatory symptoms which could be useful as biomarkers. The virome of tooth plaque was found to be significantly different between those with and without periodontal disease (113).

Conclusion

Phages have a relatively short history, having been independently discovered in 1915 and 1917, but they have been playing an important role in bacterial evolution for billions of years. Understanding phage biology is crucial to their implementation as tools. Phages have relatively small genomes, require a bacterial host for replication, and can present a number of different lifecycles. Phage taxonomy is complicated by

their lack of conserved genes and the large variations between phages. Phages and their hosts are constantly evolving for infectivity and resistance to one another with varying levels of success. Phages are currently widely used but this could increase further over time as knowledge and need increases. Community dynamics can be studied through phage-host lab experiments and metagenomics studies. Changing attitudes and rising levels of antimicrobial resistance have renewed interest in phage therapy and brought to light the use of phages in food safety. Endolysins are also an option for these issues. Faecal transplants may be a powerful tool in the treatment of CDI and other illnesses for which there is no adequate treatment. Phage display is a tool for the production of peptides and antibodies for countless purposes. Metagenomics studies may be the key to understanding the causes or successful treatment of many conditions.

References

1. **Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI.** 2012. Going viral: next-generation sequencing applied to phage populations in the human gut. Nat Rev Microbiol **10**:607-17.

2. **Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S, Metcalf JL, Ursell LK, Vázquez-Baeza Y, Van Treuren W, Hasan NA, Gibson MK, Colwell R, Dantas G, Knight R, Gilbert JA.** 2014. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science **345**:1048-1052.

3. **Gavotte L, Vavre F, Henri H, Ravallec M, Stouthamer R, Boulétreau M.** 2004. Diversity, distribution and specificity of WO phage infection in *Wolbachia* of four insect species. Insect Molecular Biology **13**:147-153.

4. **Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL.** 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. Nature Communications **8**:15955.

5. **Breitbart M, Wegley L, Leeds S, Schoenfeld T, Rohwer F.** 2004. Phage Community Dynamics in Hot Springs. Applied and Environmental Microbiology **70**:1633-1640.

6. **Wang Y, Zhang X.** 2010. Genome Analysis of Deep-Sea Thermophilic Phage D6E. Applied and Environmental Microbiology **76**:7861-7866.

7. **Borriss M, Helmke E, Hanschke R, Schweder T.** 2003. Isolation and characterization of marine psychrophilic phage-host systems from Arctic sea ice. Extremophiles **7**:377-384.

8. **Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF, Fuhrman JA, Lennon JT, Middelboe M, Suttle CA, Stock C, Wilson WH, Wommack KE, Wilhelm SW, Weitz JS.** 2016. Re-examination of the relationship between marine virus and microbial cell abundances. Nature Microbiology **1**:15024.

9. **Kim M-S, Park E-J, Roh SW, Bae J-W.** 2011. Diversity and Abundance of Single-Stranded DNA Viruses in Human Feces. Applied and Environmental Microbiology **77**:8062-8070.

10. **Shkoporov AN, Hill C.** 2019. Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. Cell Host Microbe **25**:195-209.

11. **Duckworth DH.** 1976. "Who discovered bacteriophage?". Bacteriological Reviews **40**:793-802.

12. **Twort FW.** 1915. An investigation on the nature of ultra-microscopic viruses. The Lancet **186**:1241-1243.

13. **Twort FW.** 1925. The discovery of the "bacteriophage.". The Lancet **205**:845.

14. **D'Herelle F, Smith GH.** 1926. The bacteriophage and its behavior. The Williams & Wilkins Company, Baltimore, Md.

15. **Aksyuk AA, Rossmann MG.** 2011. Bacteriophage Assembly. Viruses **3**:172-203.

16. **Fokine A, Rossmann MG.** 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage **4**:e28281.

17. **Fokine A, Zhang Z, Kanamaru S, Bowman VD, Aksyuk AA, Arisaka F, Rao VB, Rossmann MG.** 2013. The molecular architecture of the bacteriophage T4 neck. Journal of molecular biology **425**:1731-1744.

18. **Nelson D.** 2004. Phage Taxonomy: We Agree To Disagree. Journal of Bacteriology **186**:7029-7031.

19. **Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der**

Vlugt RA, Varsani A, Zerbini FM. 2017. Virus taxonomy in the age of metagenomics. Nature Reviews Microbiology **15**:161-168.

20. **Ackermann HW, Prangishvili D.** 2012. Prokaryote viruses studied by electron microscopy. Arch Virol **157**:1843-1849.

21. **Chaturongakul S, Ounjai P.** 2014. Phage–host interplay: examples from tailed phages and Gram-negative bacterial pathogens. Frontiers in Microbiology **5**:442.

22. **Bertozzi Silva J, Storms Z, Sauvageau D.** 2016. Host receptors for bacteriophage adsorption. FEMS Microbiology Letters **363**.

23. **Ross A, Ward S, Hyman P.** 2016. More is Better: Selecting for Broad Host Range Bacteriophages. Frontiers in Microbiology **7**.

24. **Yap ML, Klose T, Arisaka F, Speir JA, Veesler D, Fokine A, Rossmann MG.** 2016. Role of bacteriophage T4 baseplate in regulating assembly and infection. Proceedings of the National Academy of Sciences of the United States of America **113**:2654-2659.

25. **Székely AJ, Breitbart M.** 2016. Single-Stranded DNA Phages: From Early Molecular Biology Tools to Recent Revolutions in Environmental Microbiology. FEMS Microbiology Letters **363**:fnw027.

26. **Roux S, Krupovic M, Poulet A, Debroas D, Enault F.** 2012. Evolution and Diversity of the *Microviridae* Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. PLoS ONE **7**:e40418.

27. **Krishnamurthy SR, Janowski AB, Zhao G, Barouch D, Wang D.** 2016. Hyperexpansion of RNA Bacteriophage Diversity. PLoS Biology **14**:e1002409.

28. **Callanan J, Stockdale SR, Shkoporov A, Draper LA, Ross RP, Hill C.** 2018. RNA Phage Biology in a Metagenomic Era. Viruses **10**:386.

29. **Hatfull GF, Hendrix RW.** 2011. Bacteriophages and their Genomes. Current opinion in virology **1**:298-303.

30. **Yuan Y, Gao M.** 2017. Jumbo Bacteriophages: An Overview. Frontiers in microbiology **8**:403-403.

31. **Brüssow H, Hendrix RW.** 2002. Phage Genomics: Small Is Beautiful. Cell **108**:13-16.

32. **Weinbauer MG.** 2004. Ecology of prokaryotic viruses. FEMS Microbiology Reviews **28**:127-181.

33. **Davies EV, Winstanley C, Fothergill JL, James CE.** 2016. The role of temperate bacteriophages in bacterial infection. FEMS Microbiol Lett **363**:fnw015.

34. **Clokie MRJ, Millard AD, Letarov AV, Heaphy S.** 2011. Phages in nature. Bacteriophage **1**:31-45.

35. **Fischetti VA.** 2007. In vivo acquisition of prophage in *Streptococcus pyogenes*. Trends in Microbiology **15**:297-300.

36. **Ripp S, Miller RV.** 1997. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. Microbiology **143**:2065-2070.

37. **Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, Stotland A, Wolkowicz R, Cutting AS, Doran KS, Salamon P, Youle M, Rohwer F.** 2013. Bacteriophage adhering to mucus provide a non–host-derived immunity. Proceedings of the National Academy of Sciences **110**:10771-10776.

38. **Brüssow H, Canchaya C, Hardt W-D.** 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. Microbiology and Molecular Biology Reviews **68**:560-602.

39. **Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB, Westh H, Penadés JR, Ingmer H.** 2016. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. Nature Communications **7**:13333.

40. **Bull JJ, Vegge CS, Schmerer M, Chaudhry WN, Levin BR.** 2014. Phenotypic Resistance and the Dynamics of Bacterial Escape from Phage Control. PLoS ONE **9**:e94690.

41. **Labrie SJ, Samson JE, Moineau S.** 2010. Bacteriophage resistance mechanisms. Nature Reviews Microbiology **8**:317-327.

42. **Levin BR, Bull JJ.** 2004. Population and evolutionary dynamics of phage therapy. Nat Rev Microbiol **2**:166-73.

43. **Lu M-J, Henning U.** 1994. Superinfection exclusion by T-even-type coliphages. Trends in Microbiology **2**:137-139.

44. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science **315**:1709-1712.

45. **Dupuis M-È, Villion M, Magadán AH, Moineau S.** 2013. CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. Nature Communications **4**:3087.

46. **Chopin M-C, Chopin A, Bidnenko E.** 2005. Phage abortive infection in *lactococci*: variations on a theme. Current Opinion in Microbiology **8**:473-479.

47. **Chibani-Chennoufi S, Bruttin A, Dillmann M-L, Brüssow H.** 2004. Phage-Host Interaction: an Ecological Perspective. Journal of Bacteriology **186**:3677-3686.

48. **Lenski RE, Levin BR.** 1985. Constraints on the Coevolution of Bacteria and Virulent Phage: A Model, Some Experiments, and Predictions for Natural Communities. The American Naturalist **125**:585-602.

49. **Smith HW, Huggins MB.** 1983. Effectiveness of Phages in Treating Experimental *Escherichia coli* Diarrhoea in Calves, Piglets and Lambs. Microbiology **129**:2659-2675.

50. **Samson JE, Magadán AH, Sabri M, Moineau S.** 2013. Revenge of the phages: defeating bacterial defences. Nature Reviews Microbiology **11**:675-687.

51. **Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A.** 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol **7**:828-36.

52. **Abeles SR, Pride DT.** 2014. Molecular bases and role of viruses in the human microbiome. J Mol Biol **426**:3892-3906.

53. **Buckling A, Rainey PB.** 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society B: Biological Sciences **269**:931-936.

54. **Pal C, Macia MD, Oliver A, Schachar I, Buckling A.** 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. Nature **450**:1079-1081.

55. **Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD.** 2013. Rapid evolution of the human gut virome. Proceedings of the National Academy of Sciences **110**:12450-12455.

56. **Gama JA, Reis AM, Domingues I, Mendes-Soares H, Matos AM, Dionisio F.** 2013. Temperate Bacterial Viruses as Double-Edged Swords in Bacterial Warfare. PLoS ONE **8**:e59043.

57. **Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP.** 2013. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes **4**:4-16.

58. **Kutateladze M, Adamia R.** 2008. Phage therapy experience at the Eliava Institute. Médecine et Maladies Infectieuses **38**:426-430.

59. **Anonymous.** 1931. Limitations of bacteriophage therapy. Journal of the American Medical Association **96**:693-693.

60. **Summers WC.** 2012. The strange history of phage therapy. Bacteriophage **2**:130-133.

61. **O'Neill J.** 2014. Review on Antimicrobial Resistance Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. Review on

Antimicrobial Resistance:https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf.

62. **World Health Organization.** 2017. WHO Seventieth World Health Assembly Proposed programme budget 2018-2019.http://apps.who.int/gb/ebwha/pdf_files/WHA70/A70_7-en.pdf.

63. **World Health Organization.** 2017. Antibacterial agents in clinical development An analysis of the antibacterial clinical development pipeline, including tuberculosis. WHO reference number: WHO/EMP/IAU/20171:http://apps.who.int/iris/bitstream/handle/10665/258965/WHO-EMP-IAU-2017.11-eng.pdf?sequence=1.

64. **Robinson TP, Bu DP, Carrique-Mas J, Fèvre EM, Gilbert M, Grace D, Hay SI, Jiwakanon J, Kakkar M, Kariuki S, Laxminarayan R, Lubroth J, Magnusson U, Thi Ngoc P, Van Boeckel TP, Woolhouse MEJ.** 2016. Antibiotic resistance is the quintessential One Health issue. Transactions of The Royal Society of Tropical Medicine and Hygiene **110**:377-380.

65. **Chan BK, Abedon ST, Loc-Carrillo C.** 2013. Phage cocktails and the future of phage therapy. Future Microbiol **8**:769-83.

66. **Pelfrene E, Willebrand E, Cavaleiro Sanches A, Sebris Z, Cavaleri M.** 2016. Bacteriophage therapy: a regulatory perspective. Journal of Antimicrobial Chemotherapy **71**:2071-2074.

67. **Torres-Barceló C, Hochberg ME.** 2016. Evolutionary Rationale for Phages as Complements of Antibiotics. Trends in Microbiology **24**:249-256.

68. **LaVergne S, Hamilton T, Biswas B, Kumaraswamy M, Schooley RT, Wooten D.** 2018. Phage Therapy for a Multidrug-Resistant *Acinetobacter baumannii* Craniectomy Site Infection. Open Forum Infectious Diseases **5**:ofy064.

69. **Wright A, Hawkins CH, Änggård EE, Harper DR.** 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. Clinical Otolaryngology **34**:349-357.

70. **D'Herelle F.** 1931. Bacteriophage as a Treatment in Acute Medical and Surgical Infections. Bull N Y Acad Med **7**:329-48.

71. **Zhvania P, Hoyle NS, Nadareishvili L, Nizharadze D, Kutateladze M.** 2017. Phage Therapy in a 16-Year-Old Boy with Netherton Syndrome. Frontiers in Medicine **4**.

72. **Dufour N, Debarbieux L, Fromentin M, Ricard JD.** 2015. Treatment of Highly Virulent Extraintestinal Pathogenic *Escherichia coli* Pneumonia With Bacteriophages. Crit Care Med **43**:e190-8.

73. **Biswas B, Adhya S, Washart P, Paul B, Trostel AN, Powell B, Carlton R, Merril CR.** 2002. Bacteriophage Therapy Rescues Mice Bacteremic from a Clinical Isolate of Vancomycin-Resistant *Enterococcus faecium*. Infection and Immunity **70**:204-210.

74. **Atterbury RJ, Van Bergen MAP, Ortiz F, Lovell MA, Harris JA, De Boer A, Wagenaar JA, Allen VM, Barrow PA.** 2007. Bacteriophage Therapy To Reduce *Salmonella* Colonization of Broiler Chickens. Applied and Environmental Microbiology **73**:4543-4549.

75. **Matsuzaki S, Yasuda M, Nishikawa H, Kuroda M, Ujihara T, Shuin T, Shen Y, Jin Z, Fujimoto S, Nasimuzzaman MD, Wakiguchi H, Sugihara S, Sugiura T, Koda S, Muraoka A, Imai S.** 2003. Experimental Protection of Mice against Lethal *Staphylococcus aureus* Infection by Novel Bacteriophage ϕMR11. The Journal of Infectious Diseases **187**:613-624.

76. **Wagenaar JA, Bergen MAPV, Mueller MA, Wassenaar TM, Carlton RM.** 2005. Phage therapy reduces *Campylobacter jejuni* colonization in broilers. Veterinary Microbiology **109**:275-283.

77. **Patey O, McCallin S, Mazure H, Liddle M, Smithyman A, Dublanchet A.** 2018. Clinical Indications and Compassionate Use of Phage Therapy: Personal Experience and Literature Review with a Focus on Osteoarticular Infections. Viruses **11**:e18.

78. **Borysowski J, Weber-Dabrowska B, Gorski A.** 2006. Bacteriophage endolysins as a novel class of antibacterial agents. Exp Biol Med (Maywood) **231**:366-77.

79. **Horgan M, O'Flynn G, Garry J, Cooney J, Coffey A, Fitzgerald GF, Ross RP, McAuliffe O.** 2009. Phage lysin LysK can be truncated to its CHAP domain and retain lytic activity against live antibiotic-resistant *staphylococci*. Appl Environ Microbiol **75**:872-4.

80. **García P, Rodríguez L, Rodríguez A, Martínez B.** 2010. Food biopreservation: promising strategies using bacteriocins, bacteriophages and endolysins. Trends in Food Science & Technology **21**:373-382.

81. **Fan J, Zeng Z, Mai K, Yang Y, Feng J, Bai Y, Sun B, Xie Q, Tong Y, Ma J.** 2016. Preliminary treatment of bovine mastitis caused by *Staphylococcus*

aureus, with trx-SA1, recombinant endolysin of *S. aureus* bacteriophage IME-SA1. Veterinary Microbiology **191**:65-71.

82. **Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, Kuzeljevic B, Gold MJ, Britton HM, Lefebvre DL, Subbarao P, Mandhane P, Becker A, McNagny KM, Sears MR, Kollmann T, Mohn WW, Turvey SE, Brett Finlay B.** 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science Translational Medicine **7**:307ra152.

83. **Mayer MJ, Narbad A, Gasson MJ.** 2008. Molecular Characterization of a *Clostridium difficile* Bacteriophage and Its Cloned Biologically Active Endolysin. Journal of Bacteriology **190**:6734-6740.

84. **Bai J, Kim Y-T, Ryu S, Lee J-H.** 2016. Biocontrol and Rapid Detection of Food-borne Pathogens Using Bacteriophages and Endolysins. Frontiers in Microbiology **7**:474.

85. **Bojanova DP, Bordenstein SR.** 2016. Fecal Transplants: What Is Being Transferred? PLoS Biol **14**:e1002503.

86. **Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, McDonnell SA, Nolan JA, Sutton TDS, Dalmasso M, McCann A, Ross RP, Hill C.** 2018. Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome **6**:68.

87. **Draper LA, Ryan FJ, Smith MK, Jalanka J, Mattila E, Arkkila PA, Ross RP, Satokari R, Hill C.** 2018. Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. Microbiome **6**:220.

88. **Zuo T, Wong SH, Lam LYK, Lui R, Cheung K, Tang W, Ching J, Wu JC, Chan FK, Yu J, Sung JJ, Ng SC.** 2017. Bacteriophage Transfer during Fecal Microbiota Transplantation is Associated with Treatment Response in *Clostridium Difficile* Infection. Gastroenterology **67**:634-643.

89. **Broecker F, Klumpp J, Moelling K.** 2016. Long-term microbiota and virome in a Zürich patient after fecal transplantation against *Clostridium difficile* infection. Annals of the New York Academy of Sciences **1372**:29-41.

90. **Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, Grasis JA, Cassidy L, Tholey A, Fickenscher H, Seegert D, Rosenstiel P, Schreiber S.** 2016. Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With *Clostridium difficile* Infection. Gastroenterology **152**:799-811.

91. **Smibert OC, Guo CW, Khoo C, Thursky KA, Sandhu S, Slavin MA.** 2019. Microbiome transplantation and modulation of immune related adverse events. EClinicalMedicine **8**:10-11.

92. **Chehoud C, Dryga A, Hwang Y, Nagy-Szakal D, Hollister EB, Luna RA, Versalovic J, Kellermayer R, Bushman FD.** 2016. Transfer of Viral Communities between Human Individuals during Fecal Microbiota Transplantation. mBio **7**:e00322-16.

93. **Huttner BD, de Lastours V, Wassenberg M, Maharshak N, Mauris A, Galperine T, Zanichelli V, Kapel N, Bellanger A, Olearo F, Duval X, Armand-Lefevre L, Carmeli Y, Bonten M, Fantin B, Harbarth S, Colle L, Kloosterman F, van Bentum-Puijk W, Vlooswijk J, Andremont A, Ben Hayoun M, Canoui E, Chabrol A, Gamany N, Lafaurie M, Lefort A, Lepeule R, Louis Z, Rondinaud E, Sadou-Yaye H, Sarfati L, Zarrouk V,**

Brossier C, Carrez L, Lazarevic V, Renzi G, von Dach E, Cohen Percia S, Shvartz R, Lellouche J. 2019. A 5-day course of oral antibiotics followed by faecal transplantation to eradicate carriage of multidrug-resistant *Enterobacteriaceae*: a randomized clinical trial. Clinical Microbiology and Infection **25**:830-838.

94. **U.S. Food and Drug Administration.** 2019. Important Safety Alert Regarding Use of Fecal Microbiota for Transplantation and Risk of Serious Adverse Reactions Due to Transmission of Multi-Drug Resistant Organisms. https://wwwfdagov/vaccines-blood-biologics/safety-availability-biologics/important-safety-alert-regarding-use-fecal-microbiota-transplantation-and-risk-serious-adverse.

95. **U.S. Food and Drug Administration.** 2019. Information Pertaining to Additional Safety Protections Regarding Use of Fecal Microbiota for Transplantation – Screening and Testing of Stool Donors for Multi-drug Resistant Organisms. https://wwwfdagov/vaccines-blood-biologics/safety-availability-biologics/information-pertaining-additional-safety-protections-regarding-use-fecal-microbiota-transplantation.

96. **U.S. Food and Drug Administration.** 2019. GRAS notices. https://wwwaccessdatafdagov/scripts/fdcc/?set=GRASNotices&sort=GRN_No&order=DESC&startrow=1&type=basic&search=phage.

97. **Wu C-H, Liu IJ, Lu R-M, Wu H-C.** 2016. Advancement and applications of peptide phage display technology in biomedical science. Journal of Biomedical Science **23**:8.

98. **Brown KC.** 2010. Peptidic tumor targeting agents: the road from phage display peptide selections to clinical applications. Current pharmaceutical design **16**:1040-1054.

99. **Hu Y-F, Zhao D, Yu X-L, Hu Y-L, Li R-C, Ge M, Xu T-Q, Liu X-B, Liao H-Y.** 2017. Identification of Bacterial Surface Antigens by Screening Peptide Phage Libraries Using Whole Bacteria Cell-Purified Antisera. Frontiers in Microbiology **8**.

100. **Hairul Bahara NH, Tye GJ, Choong YS, Ong EBB, Ismail A, Lim TS.** 2013. Phage display antibodies for diagnostic applications. Biologicals **41**:209-216.

101. **Frenzel A, Schirrmann T, Hust M.** 2016. Phage display-derived human antibodies in clinical development and therapy. mAbs **8**:1177-1194.

102. **Kim K-H, Bae J-W.** 2011. Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses. Applied and Environmental Microbiology **77**:7663-7668.

103. **Castro-Mejía JL, Muhammed MK, Kot W, Neve H, Franz CMAP, Hansen LH, Vogensen FK, Nielsen DS.** 2015. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. Microbiome **3**:1-14.

104. **Roux S, Hallam SJ, Woyke T, Sullivan MB.** 2015. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. eLife **4**:e08490.

105. **Hesse U, van Heusden P, Kirby BM, Olonade I, van Zyl LJ, Trindade M.** 2017. Virome Assembly and Annotation: A Surprise in the Namib Desert. Frontiers in Microbiology **8**:13.

106. **Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA.** 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun **5**:4498.

107. **Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, Koonin EV.** 2018. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. Nature Microbiology **3**:38-46.

108. **Shkoporov A, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C.** 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. Nature Communications **9**:4781.

109. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI.** 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466**:334-338.

110. **Manrique P, Bolduc B, Walk ST, Oost J, Vos WM, Young MJ.** 2016. Healthy human gut phageome. Proc Natl Acad Sci USA **113**:10400-10405.

111. **Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DP, Keshavarzian A, Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D,**

Parkes M, Virgin HW. 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell **160**:447-60.

112. **Pérez-Brocal V, García-López R, Nos P, Beltrán B, Moret I, Moya A.** 2015. Metagenomic Analysis of Crohn's Disease Patients Identifies Changes in the Virome and Microbiome Related to Disease Status and Therapy, and Detects Potential Interactions and Biomarkers. Inflammatory Bowel Diseases **21**:2515-2532.

113. **Ly M, Abeles SR, Boehm TK, Robles-Sikisaka R, Naidu M, Santiago-Rodriguez T, Pride DT.** 2014. Altered Oral Viral Ecology in Association with Periodontal Disease. mBio **5**:e01133-14.

Tables and Figures

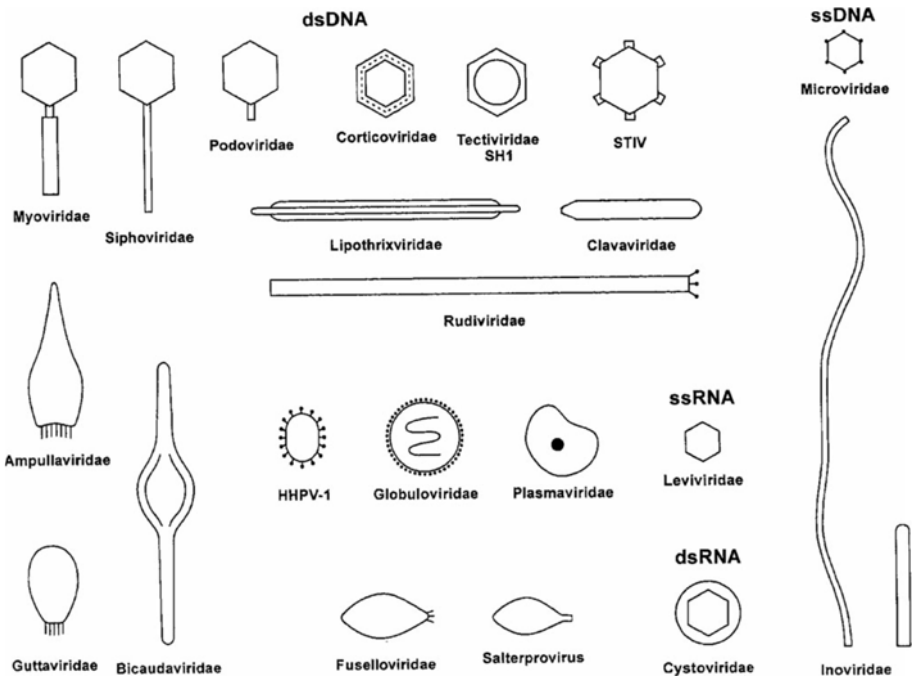


Figure 1. Morphologies of prokaryotic virus families. Figure sourced from H. W. Ackermann and D. Prangishvili (20).

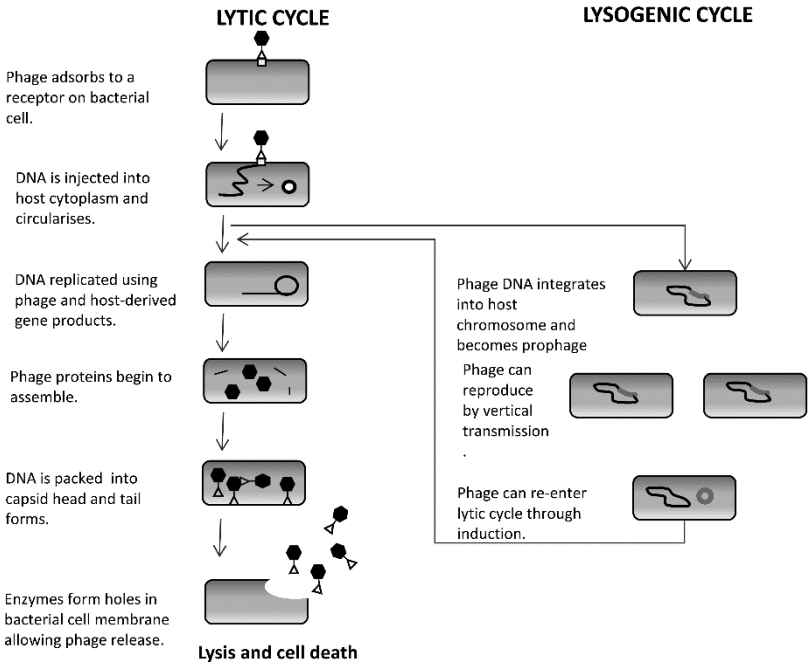


Figure 2. Replication of phages during the lytic and lysogenic cycles. In short the lytic lifecycle involves phage adsorption, DNA injection, phage replication using the host machinery, phage protein assembly, DNA packaging, and release of progeny phages. Lysogeny occurs when the phage integrates into the bacterial host DNA and replicates with the host. Prophages can also re-enter the lytic cycle. Figure sourced from Davies, *et al.* (33).

Table 1. U.S. Food and Drug Administration Generally Recognised as Safe (GRAS) Notices for phage preparations in food (96).

GRN number	Year	Target	Phage content	Notifier	Name	Food
198	2006	*L. monocytogenes*	P100	EBI Food Safety (Micreos)	Listex	Cheese
218	2007	*L. monocytogenes*	P100	EBI Food Safety (Micreos)	Listex	All foods
435	2013	*Salmonella*	SBA-1781, SKML-39, SPT-1, SSE-121, STML-13-1, STML-198	Intralytix	SalmoFresh	All foods
468	2013	*Salmonella*	FO1a, S16	Micreos	Salmonelex	Pork, poultry
528	2014	*L. monocytogenes*	LIST-36, LMSP-25, LMTA-34, LMTA-57, LMTA-94, LMTA-148	Intralytix	ListShield	All foods
603	2016	*Salmonella*	BP-63, BP-42	Phagelux	SalmoPro	All foods
672	2017	*Shigella*	SHFML-11, SHFML-26, SHSML-45, SHBML-50-1, SHSML-52-1	Intralytix	ShigaShield	All foods
724	2018	Shiga toxin-producing *E. coli*	SHFML-11, SHFML-26, SHSML-45, SHBML-50-1, SHSML-52-1	FINK TEC GmbH	*E. coli*-specific phage preparation	Beef carcasses
752	2018	*Salmonella*	BP-63, LVR16-A	Phagelux	SalmoPro (Updated from 2016 with a new phage)	
757	2018	*E. coli* O157:H7	EP, EP75	Micreos	PhageGuard	Beef
827	2019	*E. coli*	3 phage cocktail	OmniLytics		Poultry, meat, fruits, vegetables, eggs, fish, shellfish
834	2019	*E. coli*	Phages specific to shiga toxin-producing *E. coli*	Intralytix		Meat, poultry, ready-to-eat meats and poultry, fruit, vegetables, dairy products, seafood

Chapter 1b

Overcoming barriers to phage application in food and feed

Published as:

Lewis R, Hill C. 2020. Overcoming barriers to phage application in food and feed. Current Opinion in Biotechnology 61:38-44.

Abstract

Bacteriophages (phages) can play a useful role as narrow spectrum antimicrobials in food safety and in food production. Consumer attitudes towards traditional additives have led to a search for natural, potentially clean label, alternatives. At the same time, the rise in antimicrobial resistance has created a need for alternative antimicrobials for disease prevention and treatment in animal husbandry. Phages represent a viable option for both of these applications. We highlight important barriers which should be considered to improve the chance of a positive outcome when using phages in food and food production. These include the feasibility of adding high concentrations of phages, the physico-chemical properties of the food or target, how and when phages are applied, and which phages are chosen.

Introduction

Bacteriophages (phages) are bacterial viruses and use bacteria to produce new phages. Phages can replicate using lytic or lysogenic lifecycles. Lytic phages replicate by attaching to a host cell, injecting phage nucleic acid, replicating the phage nucleic acid, assembling proteins, packaging nucleic acid with head and tail proteins, and lysing the host cell to release new phages (1). Lysogenic phages integrate into the host genome and replicate along with the host. Interest is growing in the use of lytic phages as biocontrol agents of foodborne pathogens in food and food production (2). Phages can be used in a number of ways in food and feed and can also be applied as pre or post-harvest interventions. Recently there has been a change in consumer attitudes towards how food should be processed and preserved with rising interest in the concept of “natural foods” and “naturalness” (3). The importance of consumers’ perception

should not be overlooked as it has a significant effect on willingness to buy a product and can even affect how individuals enjoy and experience that product (4). The importance of this changing consumer attitude has led to the introduction of the term "clean label"; these products aim to have simple ingredient lists free from "chemical-sounding" terms and negatively perceived ingredients (5•). It has been found that consumers are more willing to accept additives from natural sources than the chemical additives that have been used for decades. This rise in clean label food has increased the demand for new methods to ensure food safety. Another reason for introducing phages is the problem of antimicrobial resistance. A WHO report on antimicrobial resistance drew attention to the contribution of antibiotic use in animals to antimicrobial resistance in general (6). Antimicrobial resistance has been termed a One Health problem and requires an integrative approach targeting the environment and the health of humans and animals (7). Historically the antibiotics used in treating livestock have been the same as, or closely related to, those used in human medicine and the use of long-term sub-therapeutic doses of antibiotics, as they have been used in livestock as growth promoters, can increase the risk of emergence of antibiotic resistance.

Phage concentration

The success of phages for use in food and feed depends on the overcoming of a number of barriers. The number of phages used is of great importance; in general, the higher the concentration of phages the more significant the reductions in target bacteria. Bai, *et al*. (8••) found that a phage cocktail at multiplicity of infections (MOIs) of 10^3 and 10^4 significantly reduced the *Salmonella* Typhimurium load of cucumbers and lettuce. The cocktail at an MOI of 10^4 led to a greater and more

sustained reduction. Phage at MOIs of 1, 10 and 10^2 significantly reduced *S.* Typhimurium on lettuce with the reductions increasing as the phage concentration increased (9•). This was also evident using SalmoFresh™, an FDA approved Salmonella lytic bacteriophage preparation, on chicken fillets (10). SalmoFresh™ applied at MOIs of 10^2 and 10^3 both significantly reduced a combination of *S.* Typhimurium, *S.* Heidelberg and *S.* Enteritidis at an initial concentration of 3 log CFU/g on chicken fillets. SalmoFresh™ at an MOI of 10^3 was significantly better than phage at an MOI of 10^2. In other examples, a single phage was used to control *Escherichia coli* on raw and cooked beef (11), while ShigaShield™ reduced Shigella levels in smoked salmon and yogurt (12••). In both cases the reduction was concentration dependent with the greatest reduction occurring at the highest phage titre tested. Using Listex™ P100, a commercially available phage against *Listeria monocytogenes*, a greater reduction was seen in tuna using an MOI of 10^2 than an MOI of 0.1 (13). The greatest effect was seen with a lower starting concentration of *L. monocytogenes* combined with a higher concentration of Listex™ P100. However, phage concentration is not a definitive indicator of a positive outcome. A low titre phage cocktail decreased chickens' mortality and morbidity in a natural outbreak of Avian Pathogenic *E. coli* (APEC) but not in experimentally infected chickens (14). This could have been due to the non-synchronous nature of the natural outbreak. In a natural outbreak all of the cases would be not at the same point in their infection cycle and the low titre phage cocktail may have controlled the infection at the early stages and interrupted the transmission. This would not have occurred in the experimental infection as all of the infections occurred at the same time.

Feasibility of adding high phage concentrations

The need for high numbers of phages for successful phage treatment brings up the question of the feasibility of generating phages at high enough titres to deploy in food or feed. *Salmonella* on post-chill chicken carcases has been suggested as a target as numbers do not usually exceed 3 log CFU/g (15). The inoculation of 25 g samples of turkey fillet with 500 µl of a 10^9 PFU/ml bacteriophage preparation was sufficient to significantly reduce *S.* Heidelberg over a seven day period and represented a reasonable volume of phage preparation to be applied. In a study of the application of Finalyse™ hidewash, ~$3x10^{10}$ phage/head of cattle in one gallon of water was used (16). This was insufficient to reduce *E. coli* O157:H7 contamination of cattle hides and carcasses but given the one gallon quantity needed it may be difficult and cost prohibitive to increase the titre used. Recombinantly expressed endolysins isolated from phages could possibly be used in place of whole phages to avoid this issue (17), though once again the economic viability of this solution would have to be investigated. Issues with GMO legislative approval and consumer acceptance may limit the use of endolysins in food safety (18).

Properties of the target matrix

The properties of the targeted food or animal in question can be a barrier to phage success. The target matrix can affect the efficiency of phages. In a study of a single phage to reduce *S.* Typhimurium in whole milk, skimmed milk, energy drink, apple juice, and liquid egg the smallest effect of the phage was seen in liquid egg (19). Phage titre increased in all foods except for liquid egg in which a decrease occurred. The highly viscous matrix of egg limiting diffusion and homogeneous distribution of

the phage particles was suggested as a reason for this reduced activity and the decrease in phage numbers. Similarly when looking at a phage cocktail to reduce *S.* Enteritidis in milk, cabbage, and chicken breast the greatest effect was seen in milk (20). Again it was proposed that the liquid allowed greater diffusion of the phages. Using a single phage it was necessary to use an MOI of 10^5 to reduce *S.* Typhimurium load in liquid egg and fruit juice while an MOI of 10^7 was required to give a similar reduction in cooked beef and chicken (21). Salmonelex™, a commercially available phage cocktail, was effective in reducing *Salmonella* on chicken (22). Salmonelex™ diluted in tap water was more effective than in filtered tap water with reduced calcium, magnesium, and sodium. Temperature can also affect the activity of phages. The antibacterial activity of ListShield™, a commercially available phage against *L. monocytogenes*, was reduced at 12˚C compared to 4˚C (23). When using phages to reduce *Staphylococcus aureus* during the manufacturing of cheese it was observed that phage titre decreased with decreasing pH (24). Antacids can be given before or in combination with phages in animal application to improve the survival of phages in the gut by increasing pH (25, 26).

Application method of phages

The method of phage application can also be a barrier to a positive outcome. Although dipping and spraying are common methods for phage application, they can have negative effects (27). Dipping and spraying can release phage particles into the environment, dipping liquid can be a source of cross contamination and spraying equipment may not be available in the processing environment. Different methods can be used to suit the situation or to expand the conditions of activity of the phages.

Microencapsulated phages significantly reduced *E. coli* on tomatoes and maintained the reduction for 5 days (28). Microencapsulation reduced the UV sensitivity of the phages and increased the survival of phages at pH 3-7 and extreme temperatures. Free phages, paper impregnated with phages, and encapsulated phages all immediately reduced *E. coli* on alfalfa seeds and sprouts (27). After five days free phages and encapsulated phages had a significant effect compared to untreated controls. This can be case dependent as in the same study free phages reduced *L. monocytogenes* by 3 log CFU/g in cantaloupes while encapsulated phages reduced numbers by only 1 log CFU/g over the same period. This was suggested to be due to how the bacteria attached to different food matrices represented by cantaloupe and alfalfa sprouts. A phage cocktail applied to chicken feed was as effective as phage introduced by crop gavage and represents a much easier mode of inoculation (25). Phages can also be applied in more unexpected ways such as in the depuration of bivalves where phages were added to the depuration tank water (29). Longer treatment time was required in control tanks to obtain comparable reductions to those achieved when using phage tanks.

When phages should be applied

Phage interventions must be carried out at the correct stage of processing for a positive outcome and to avoid reintroduction of bacteria. For example, cucumbers that were sprayed with SalmoFresh™ and then sliced did not show a significant reduction in *S.* Newport while unsliced cucumbers did show a significant reduction (30). It is believed that insufficient numbers of phages were transferred by cutting to achieve a significant reduction. A cocktail of six phages reduced the symptoms of *Pectobacterium atrosepticum* soft rot in potato tubers significantly (31•). However,

phage alone increased the disease severity in uninfected control tubers. This was suggested to be caused by the presence of enzymes or metabolites from phage production. This could limit the use of phages as a prophylactic treatment but not as an intervention as the cost to uninfected tubers could outweigh the benefit of reduction of symptoms in uninfected tubers. Regarding the issue of reinfection pigs can become infected with *Salmonella* after just two hours in a contaminated abattoir environment so would need to be treated at the correct time to avoid this (32). To reduce *Campylobacter* in chicken, interventions should be performed at multiple stages as *Campylobacter* can re-enter the food due to its ubiquitous nature on farms (33).

Phage selection

There are a number of disadvantages associated with the use of phages but this barrier can be overcome by careful selection of phages. The emergence of phage resistant bacteria is a risk (34). To decrease the issue of resistance treatments can be rotated or cocktails of phages can be used instead of single phages. Phages should only be applied when there is no risk of treated bacteria being reintroduced into the processing environment and causing issues with resistance. The highly specific nature of phages to their host has historically been viewed as a disadvantage in that they may only infect a limited number of strains (35). This view is changing as it is recognised that their specificity limits negative effects to the surrounding bacterial community. For example, it would be possible to use phages to reduce enteropathogenic or Shiga toxigenic *E. coli* during milk fermentation without compromising the performance of starter cultures (36). Cocktails of multiple phages can also be used to increase the number of strains targeted (37•). Phages can carry virulence genes or antibiotic

resistance genes but this can be largely avoided by using lytic rather than lysogenic phages (38). There are also concerns about the immunogenicity of phages and the cytotoxicity which could result by lysis of target bacteria. Phages have been found to cause no adverse reactions in rats and mice suggesting they are safe for human use (39, 40). Phages against *Staphylococcus aureus* in cheddar manufacture did not increase enterotoxin production (41). Phages are constantly encountered by humans since they are found naturally in the gut of humans (42), on their skin (43), in animals (44), in sewage treatment systems (45), and during the breakdown of food fermentations (46) among others.

Conclusion

Phages show promise for use in controlling bacterial pathogens as additives from natural sources, which may be more readily accepted by consumers than traditional additives, and also in the growing fight against antibiotic resistance. Phages may only reduce bacterial populations in food and not eliminate them completely, but this is not a serious issue. Criteria are put in place by groups, such the European Union, for acceptable levels of pathogens in food depending on the pathogen, food, and intended consumer with some required to be absent and some acceptable at low levels (47). There is zero tolerance for *Salmonella* in foods such as pre-cut vegetables, *L. monocytogenes* in food for infants and for medical purposes, and *E. coli* O157:H7 on sprouts. Phages reduced *S.* Typhimurium in whole milk and skimmed milk to undetectable, and therefore authorised, levels (19). Phages reduced *L. monocytogenes* to undetectable levels on tuna and Spanish dry cured ham (13, 23). Coagulase-positive

staphylococci are permitted in cheeses during manufacture up to 10^4 CFU/g. Phages have been successful in reducing *S. aureus* in cheddar cheese to safe levels (41).

Bacteria use a number of methods to defend against phage infection at various points in the infection process (48). Bacteria can inhibit phage adsorption by altering or blocking receptors. Superinfection exclusion occurs when a prophage causes the expression of proteins which stops phage injection. Restriction modification methylates host DNA and cleaves invading unmethylated DNA. Bacteria using the CRISPR-Cas system (clustered, regularly interspaced, short palindromic repeat) integrate small fragments of invading DNA and cleave DNA with this sequence. Abortive infection limits the spread of the phage by the death of infected host cells. Care must be taken in the application of phages and their use to ensure they achieve a favourable outcome before they become a widely used and accepted aid in food processing. Phage concentration, the feasibility of adding a high concentration of phages, the properties of the food or animal to be treated, how phages are applied, when they are applied, and what phages are used are all barriers which must be overcome when designing a process using phages for decontamination or bacterial control.

References

1. **Sulakvelidze A, Alavidze Z, Morris JG, Jr.** 2001. Bacteriophage therapy. Antimicrobial agents and chemotherapy **45:**649-659.

2. **Stone E, Campbell K, Grant I, McAuliffe O.** 2019. Understanding and Exploiting Phage-Host Interactions. Viruses **11:**567.

3. **Román S, Sánchez-Siles LM, Siegrist M.** 2017. The importance of food naturalness for consumers: Results of a systematic review. Trends in Food Science & Technology **67:**44-57.

4. **Naanwaab C, Yeboah O-A, Ofori Kyei F, Sulakvelidze A, Goktepe I.** 2014. Evaluation of consumers' perception and willingness to pay for bacteriophage treated fresh produce. Bacteriophage **4:**e979662.

5. • **Aschemann-Witzel J, Varela P, Peschel AO.** 2019. Consumers' categorization of food ingredients: Do consumers perceive them as 'clean label' producers expect? An exploration with projective mapping. Food Quality and Preference **71:**117-128.

The study explored consumers' perception and categorisation of food ingredients. Unknown ingredients were perceived negatively and there was a preference for natural ingredients.

6. **World Health Organisation.** 2017. Antibacterial agents in clinical development An analysis of the antibacterial clinical development pipeline, including tuberculosis. WHO reference number: WHO/EMP/IAU/2017.1 2017:http://apps.who.int/iris/bitstream/handle/10665/258965/WHO-EMP-IAU-252017.258911-eng.pdf?sequence=258961.

7. **Walsh TR.** 2018. A one-health approach to antimicrobial resistance. Nature Microbiology **3:**854-855.

8. •• **Bai J, Jeon B, Ryu S.** 2019. Effective inhibition of *Salmonella* Typhimurium in fresh produce by a phage cocktail targeting multiple host receptors. Food Microbiology **77:**52-60.

The authors isolated and developed a phage cocktail effective against *Salmonella* species and highlighted the importance of choosing phages free from genes involved in lysogeny and virulence factor production.

9. • **Huang C, Shi J, Ma W, Li Z, Wang J, Li J, Wang X.** 2018. Isolation, characterization, and application of a novel specific *Salmonella* bacteriophage in different food matrices. Food Research International **111:**631-641.

A phage was tested in a range of foods, temperatures and concentrations. Greater bacterial reduction by phage occurred at higher temperatures as bacteria were more actively growing.

10. **Sukumaran AT, Nannapaneni R, Kiess A, Sharma CS.** 2015. Reduction of *Salmonella* on chicken meat and chicken skin by combined or sequential application of lytic bacteriophage with chemical antimicrobials. International Journal of Food Microbiology **207:**8-15.

11. **Hudson J, Billington C, Wilson T, On S.** 2015. Effect of phage and host concentration on the inactivation of *Escherichia coli* O157:H7 on cooked and raw beef. Food Science and Technology International **21:**104-109.

12. •• **Soffer N, Woolston J, Li M, Das C, Sulakvelidze A.** 2017. Bacteriophage preparation lytic for *Shigella* significantly reduces *Shigella sonnei* contamination in various foods. PLOS ONE **12:**e0175256.

This study looked at the use of ShigaShield™, a commercially available phage cocktail, in deli meat, smoked salmon, pre-cooked chicken, lettuce, melon and yogurt at various concentrations. virulent and/or toxic genes) present. No toxin genes, virulence genes, repressor genes, integrases, recombinases or any bacterial gene listed

in the US Code for Federal Regulations (40 CFR §725.421) were detected in the genomes of the phages.

13. **Migueis S, Saraiva C, Esteves A.** 2017. Efficacy of LISTEX P100 at Different Concentrations for Reduction of *Listeria monocytogenes* Inoculated in Sashimi. J Food Prot **80:**2094-2098.

14. **Oliveira A, Sereno R, Azeredo J.** 2010. In vivo efficiency evaluation of a phage cocktail in controlling severe colibacillosis in confined conditions and experimental poultry houses. Veterinary Microbiology **146:**303-308.

15. **Sharma CS, Dhakal J, Nannapaneni R.** 2015. Efficacy of Lytic Bacteriophage Preparation in Reducing *Salmonella* In Vitro, on Turkey Breast Cutlets, and on Ground Turkey. Journal of Food Protection **78:**1357-1362.

16. **Arthur TM, Kalchayanand N, Agga GE, Wheeler TL, Koohmaraie M.** 2017. Evaluation of Bacteriophage Application to Cattle in Lairage at Beef Processing Plants to Reduce *Escherichia coli* O157:H7 Prevalence on Hides and Carcasses. Foodborne Pathogens and Disease **14:**17-22.

17. **Fan J, Zeng Z, Mai K, Yang Y, Feng J, Bai Y, Sun B, Xie Q, Tong Y, Ma J.** 2016. Preliminary treatment of bovine mastitis caused by *Staphylococcus aureus*, with trx-SA1, recombinant endolysin of *S. aureus* bacteriophage IME-SA1. Veterinary Microbiology **191:**65-71.

18. **Misiou O, van Nassau TJ, Lenz CA, Vogel RF.** 2018. The preservation of *Listeria*-critical foods by a combination of endolysin and high hydrostatic pressure. International Journal of Food Microbiology **266:**355-362.

19. **Zinno P, Devirgiliis C, Ercolini D, Ongeng D, Mauriello G.** 2014. Bacteriophage P22 to challenge *Salmonella* in foods. Int J Food Microbiol **191:**69-74.

20. **Bao H, Zhang P, Zhang H, Zhou Y, Zhang L, Wang R.** 2015. Bio-Control of *Salmonella* Enteritidis in Foods Using Bacteriophages. Viruses **7:**4836-4853.

21. **Thung TY, Krishanthi Jayarukshi Kumari Premarathne JM, San Chang W, Loo YY, Chin YZ, Kuan CH, Tan CW, Basri DF, Jasimah Wan Mohamed Radzi CW, Radu S.** 2017. Use of a lytic bacteriophage to control *Salmonella* Enteritidis in retail food. LWT **78:**222-225.

22. **Grant AQ, Parveen S, Schwarz J, Hashem F, Vimini B.** 2017. Reduction of *Salmonella* in ground chicken using a bacteriophage. Poultry Science **96:**2845-2852.

23. **Gutiérrez D, Rodríguez-Rubio L, Fernández L, Martínez B, Rodríguez A, García P.** 2017. Applicability of commercial phage-based products against *Listeria monocytogenes* for improvement of food safety in Spanish dry-cured ham and food contact surfaces. Food Control **73:**1474-1482.

24. **Bueno E, García P, Martínez B, Rodríguez A.** 2012. Phage inactivation of *Staphylococcus aureus* in fresh and hard-type cheeses. International Journal of Food Microbiology **158:**23-27.

25. **Carvalho CM, Gannon BW, Halfhide DE, Santos SB, Hayes CM, Roe JM, Azeredo J.** 2010. The in vivo efficacy of two administration routes of a phage

cocktail to reduce numbers of *Campylobacter coli* and *Campylobacter jejuni* in chickens. BMC Microbiology **10:**232.

26. **Hammerl JA, Jäckel C, Alter T, Janzcyk P, Stingl K, Knüver MT, Hertwig S.** 2014. Reduction of *Campylobacter jejuni* in Broiler Chicken by Successive Application of Group II and Group III Phages. PLOS ONE **9:**e114785.

27. **Lone A, Anany H, Hakeem M, Aguis L, Avdjian A-C, Bouget M, Atashi A, Brovko L, Rochefort D, Griffiths MW.** 2016. Development of prototypes of bioactive packaging materials based on immobilized bacteriophages for control of growth of bacterial pathogens in foods. International Journal of Food Microbiology **217:**49-58.

28. **Ramirez K, Cazarez-Montoya C, Lopez-Moreno HS, Castro-del Campo N.** 2018. Bacteriophage cocktail for biocontrol of *Escherichia coli* O157:H7: Stability and potential allergenicity study. PLOS ONE **13:**e0195023.

29. **Pereira C, Moreirinha C, Teles L, Rocha RJM, Calado R, Romalde JL, Nunes ML, Almeida A.** 2017. Application of phage therapy during bivalve depuration improves *Escherichia coli* decontamination. Food Microbiology **61:**102-112.

30. **Sharma M, Dashiell G, Handy ET, East C, Reynnells R, White C, Nyarko E, Micallef S, Hashem F, Millner PD.** 2017. Survival of *Salmonella* Newport on Whole and Fresh-Cut Cucumbers Treated with Lytic Bacteriophages. Journal of Food Protection **80:**668-673.

31. **Carstens AB, Djurhuus AM, Kot W, Hansen LH.** 2019. A novel six-phage cocktail reduces *Pectobacterium atrosepticum* soft rot infection in potato tubers under simulated storage conditions. FEMS Microbiology Letters **366**

A phage cocktail was developed against *Pectobacterium atrosepticum*, the causitive agent of potato soft rot, and was effective but had a negative effect on uninfected controls.

32. **Walia K, Argüello H, Lynch H, Grant J, Leonard FC, Lawlor PG, Gardiner GE, Duffy G.** 2017. The efficacy of different cleaning and disinfection procedures to reduce *Salmonella* and *Enterobacteriaceae* in the lairage environment of a pig abattoir. International Journal of Food Microbiology **246:**64-71.

33. **Fischer S, Kittler S, Klein G, Glünder G.** 2013. Impact of a Single Phage and a Phage Cocktail Application in Broilers on Reduction of *Campylobacter jejuni* and Development of Resistance. PLOS ONE **8:**e78543.

34. **Guenther S, Herzig O, Fieseler L, Klumpp J, Loessner MJ.** 2012. Biocontrol of *Salmonella* Typhimurium in RTE foods with the virulent bacteriophage FO1-E2. International Journal of Food Microbiology **154:**66-72.

35. **Cieplak T, Soffer N, Sulakvelidze A, Nielsen DS.** 2018. A bacteriophage cocktail targeting *Escherichia coli* reduces *E. coli* in simulated gut conditions, while preserving a non-targeted representative commensal normal microbiota. Gut Microbes **9:**391-399.

36. **Tomat D, Mercanti D, Balague C, Quiberoni A.** 2013. Phage biocontrol of enteropathogenic and Shiga toxin-producing *Escherichia coli* during milk fermentation. Lett Appl Microbiol **57:**3-10.

37. **Costa P, Pereira C, Gomes ATPC, Almeida A.** 2019. Efficiency of Single
- Phage Suspensions and Phage Cocktail in the Inactivation of *Escherichia coli* and *Salmonella* Typhimurium: An In Vitro Preliminary Study. Microorganisms **7:**94.

Although about hospital acquired infections the outcome can be applied to food settings. A phage cocktail was more effective against *S.* Typhimurium than a single phage. It was determined that cocktails could be used as broad spectrum agents before a target was identified.

38. **Cui Z, Guo X, Dong K, Zhang Y, Li Q, Zhu Y, Zeng L, Tang R, Li L.** 2017. Safety assessment of *Staphylococcus* phages of the family *Myoviridae* based on complete genome sequences. Scientific Reports **7:**41259.

39. **Kang H-W, Kim J-W, Jung T-S, Woo G-J.** 2013. wksl3, a New Biocontrol Agent for *Salmonella enterica* Serovars Enteritidis and Typhimurium in Foods: Characterization, Application, Sequence Analysis, and Oral Acute Toxicity Study. Applied and Environmental Microbiology **79:**1956-1968.

40. **EFSA Panel on Biological Hazards.** 2016. Evaluation of the safety and efficacy of Listex™ P100 for reduction of pathogens on different ready-to-eat (RTE) food products. EFSA Journal **14:**e04565.

41. **El Haddad L, Roy J-P, Khalil GE, St-Gelais D, Champagne CP, Labrie S, Moineau S.** 2016. Efficacy of two *Staphylococcus aureus* phage cocktails in cheese production. International Journal of Food Microbiology **217:**7-13.

42. **Shkoporov A, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C.** 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. Nature Communications **9:**4781.

43. **van Zyl LJ, Abrahams Y, Stander EA, Kirby-McCollough B, Jourdain R, Clavaud C, Breton L, Trindade M.** 2018. Novel phages of healthy skin metaviromes from South Africa. Scientific Reports **8:**12265.

44. **Morozova V, Kozlova Y, Shedko E, Babkin I, Kurilshikov A, Bokovaya O, Bardashova A, Yunusova A, Tikunov A, Tupikin A, Ushakova T, Ryabchikova E, Tikunova N.** 2018. Isolation and characterization of a group of new *Proteus* bacteriophages. Archives of Virology **163:**2189-2197.

45. **Fernandez-Cassi X, Timoneda N, Martínez-Puchol S, Rusiñol M, Rodriguez-Manzano J, Figuerola N, Bofill-Mas S, Abril JF, Girones R.** 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Science of The Total Environment **618:**870-880.

46. **Halter MC, Zahn JA.** 2018. Characterization of a novel lytic bacteriophage from an industrial *Escherichia coli* fermentation process and elimination of virulence using a heterologous CRISPR-Cas9 system. J Ind Microbiol Biotechnol **45:**153-163.

47. **European Commission.** 2005. Commission Regulation (EC) No 2073/2005.

48. **Azam AH, Tanji Y.** 2019. Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. Applied Microbiology and Biotechnology **103:**2121-2131.

Tables and figures

Table 1. The application of phages to eradicate target microbes. In the case of ListShield™ against *L. monocytogenes* in Spanish dry cured ham bacterial numbers increased after treatment at 12˚C but decreased after treatment at 4˚C, this increase was represented by a (+) (23).

Phage cocktail	Target organism	Food matrix	MOI	Application method	Bacterial reduction	Reference
BSPM4, BSP101, BSP22A	*S.* Typhimurium	Iceberg lettuce	10^3	Applied to surface	1.1-1.9 log CFU/cm^2	(8)
			10^4		2.8-3.9 log CFU/cm^2	
		Cucumber	10^3		0.7-1.2 log CFU/cm^2	
			10^4		2.5-2.8 log CFU/cm^2	
LPST10	*S.* Typhimurium	Lettuce	1	Applied to surface	0.7-1.7 $\log_{10}$ CFU/cm^2	(9)
			10		1.1-1.7 $\log_{10}$ CFU/cm^2	
			10^2		1.9-2.7 $\log_{10}$ CFU/cm^2	
SalmoFresh™	*S.* Typhimurium, *S.* Heidelberg, *S.* Enteritidis	Chicken breast fillets	10^2	Applied to surface	0.6 log CFU/g	(10)
			10^3		1.1 log CFU/g	
FAHEc1	*E. coli*	Raw beef	1	Applied to surface	1 log CFU/cm^2	(11)
			$5X10^2$		2.7 log CFU/cm^2	
		Cooked beef	10^2		1 log CFU/cm^2	
			10^4		5 log CFU/cm^2	
ShigaShield™	*Shigella sonnei*	Smoked salmon	$2.25X10^2$	Spraying	0.16 log CFU/g	(12)
			$2.25X10^3$		0.50 log CFU/g	
			$2.25X10^4$		1.098 log CFU/g	
		Yogurt	$4.5X10^2$	Mixed in to food	0.07 log CFU/g	
			$4.5X10^3$		0.26 log CFU/g	
			$4.5X10^4$		1.01 log CFU/g	
Listex™ P100	*L. monocytogenes*	Tuna sashimi	0.1	Applied to surface	0.62 log CFU/g	(13)
			10^2		1.11 log CFU/g	
			10^3		1.08 log CFU/g	
			10^6		2.35 log CFU/g	
phi F78E	Avian pathogenic *E. coli*	Experimentally infected chickens	30	Oral gavage and spraying	No reduction in morbidity or mortality	(14)

		Naturally infected chickens		Oral gavage and spraying	Decreased mortality to below 0.5% in no more than 3 weeks	
SalmoFresh™	*S.* Heidelberg	Turkey breast	10^4	Applied to surface	1.3 log CFU/g	(15)
Finalyse™	*E. coli*	Live cattle	~$3x10^{10}$ phage/head of cattle in one gallon of water. Looked at prevalence in cattle not bacterial numbers.	Applied to hide surface	6.6 % reduction in prevalence	(16)
P22	*S.* Typhimurium	Whole milk Skimmed milk Energy drink Apple juice Liquid egg	10^8	Mixed in to liquid	4.45 log CFU/ml 4.32 log CFU/ml 2.09 log CFU/ml 2.06 log CFU/ml 0.96 log CFU/ml	(19)
PA13076, PC2184	*S.* Enteritidis	Milk Cabbage Chicken breast	10^4	Applied to surface Applied to surface Mixed in to liquid	4 log CFU/sample 3.86 log CFU/sample 2.5 log CFU/sample	(20)
SE07	*S.* Enteritidis	Liquid egg Fruit juice Beef Chicken	10^5 10^5 10^7 10^7	Mixed in to liquid Mixed in to liquid Spraying Spraying	1.96 log CFU/ml 2.06 log CFU/ml 2.03 log CFU/ml 2.18 log CFU/ml	(21)
Salmonelex™	*S.* Newport, *S.* Typhimurium, *S.* Heidelberg, *S.* Enteritidis	Skinless chicken legs and thighs	10^3	Applied to surface in tap water Applied to surface in filtered water	0.39 log CFU/cm^2 0.23 log CFU/cm^2	(22)
ListShield™	*L. monocytogenes*	Spanish dry cured ham	10^2 10^2 10^3 10^3 10^4	Applied to surface 4˚C Applied to surface 12˚C Applied to surface 4˚C Applied to surface 12˚C Applied to surface 4˚C	1.5 log CFU/cm^2 +1.5 log CFU/cm^2 2 log CFU/cm^2 +3 log CFU/cm^2 3.5 log CFU/cm^2	(23)

vB_SauS-phi-IPLA35, vB_SauS-phi-SauS-IPLA88)	*S. aureus*	Fresh cheese	6	Added during cheese manufacture	3.83 log CFU/g	(24)
phiCcoIBB35, phiCcoIBB37, phiCcoIBB12	*C. jejuni*	Live chickens	15	In feed Oral gavage	1.96 log CFU/g 1.69 log CFU/g	(25)
CP68, CP14	*C. jejuni*	Live chickens	10	Oral gavage	3 log CFU/g	(26)
LinM-AG8, LmoM-AG13, LmoM-AG20	*L. monocytogenes*	Cantaloupe	10^5 $10\text{-}10^2$	Applied to surface Encapsulated phage	3 log CFU/g 1 log CFU/g	(27)
EcoM-HG2, EcoM-HG7, EcoM-HG8	*E. coli*	Alfalfa seeds and sprouts	10^4 $10\text{-}10^2$ N/A	Applied to surface Encapsulated phage Impregnated paper	1.5 log CFU/g 1.3 log CFU/g 0.6 log CFU/g	
FJLA23, FKP26, FC119, FE142	*E. coli*	Tomatoes	10^4	Spraying microencapsulated phage	2.5 log CFU/tomato	(28)
phT4A, ECA2	*E. coli*	Cockles	1	Added to depuration tank water	0.6 log CFU/g	(29)
SalmoFresh™	*S.* Newport	Cucumbers	$7.9X10^4$	Sprayed before slicing Sprayed unsliced	1 log CFU/sample 1.83 log CFU/sample	(30)
Nepra, Lelidair, Nobby, Slant, Gaspode, Momine	*Pectobacterium atrosepticum*	Potatoes	10^2	Phage wash	Reduction in disease incidence 61.3%	(31)
DT6	*E. coli*	Milk	$2.4X10^4$	Added during milk fermentation	1.1 log CFU/ml	(36)
wksl3	*S.* Enteritidis	Chicken skin	$5X10^3$	Spraying	2.43 log CFU/cm^2	(39)
Team1, P68, LH1-MUT	*S. aureus*	Cheddar cheese	150	Added during cheese manufacture	2 log CFU/ml	(41)

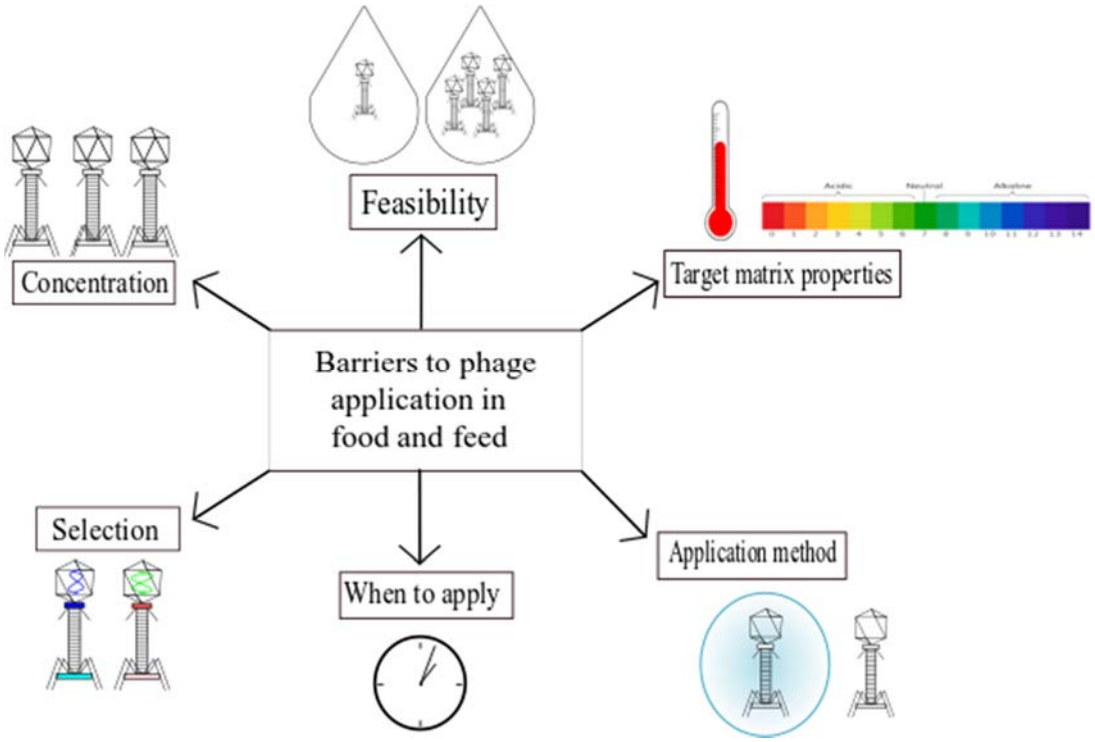


Figure 1. Graphical abstract

Chapter 2

The Effect of a Commercially Available Bacteriophage and Bacteriocin on *Listeria monocytogenes* in Coleslaw

Published as:

Lewis R, Bolocan AS, Draper LA, Ross RP, Hill C. 2019. The Effect of a Commercially Available Bacteriophage and Bacteriocin on *Listeria monocytogenes* in Coleslaw. Viruses 11.

Abstract

Changing consumer attitudes show an increased interest in non-chemical antimicrobials in food preservation and safety. This greater interest of consumers in more 'natural' or 'clean-label' food interventions is complicated by concurrent demands for minimally processed, ready-to-eat (RTE) foods with long shelf lives. Two viable interventions are bacteriophage (phage) and bacteriocins, a number of which have already been approved for use in food safety. Listeriosis is a serious foodborne infection which affects at-risk members of the population. Listeriosis incidence has increased between 2008 and 2015 and has a case fatality rate of up to 20% with antibiotic intervention. Here, we tested an intervention to attempt to control a pathogenic *Listeria monocytogenes* strain in a food model using two of these alternative antimicrobials. Phage P100 on its own had a significant effect on *L. monocytogenes* ScottA numbers in coleslaw over a 10-day period at 4˚C ($P \leq 0.001$). A combination of P100 and Nisaplin® (a commercial formulation of the lantibiotic bacteriocin, nisin) had a significant effect on the pathogen ($P \leq 0.001$). P100 and Nisaplin® in combination were more effective than Nisaplin® alone, but not P100 alone.

Introduction

Listeria monocytogenes is a small (0.5–2 μm in length), Gram positive rod-shaped bacteria, and the causative agent of listeriosis (1-3). Listeriosis is a serious foodborne illness which can occur as sporadic cases or as outbreaks, and most commonly affects the most susceptible members of the population; namely, neonates, the immunocompromised, the elderly, and pregnant women. The majority of listeriosis

cases manifest one of three clinical syndromes: Maternofetal or neonatal listeriosis, blood stream infection, or meningoencephalitis. *Listeria* is of significant concern, despite the fact that fewer than 2300 cases were reported in the European Union (EU)/European Economic Area (EEA) per year between 2008 and 2015 (4). This concern is due to an overall annual case fatality rate of 12-20%, the involvement of at risk groups, and an increase in confirmed cases between 2008 and 2015. *L. monocytogenes* has been associated with a wide range of foods such as meat, seafood, vegetables, egg products, dairy products, and ready-to-eat (RTE) foods (5, 6). Interest is growing in biopreservation: The use of the anti-microbial properties of micro-organisms and their metabolites for the preservation of food (7). This is, in part, due to the growing popularity of ready-to-eat (RTE) and minimally processed foods, which often carry an increased risk of microbiological contamination and are usually not cooked before consumption. Another reason for the rising interest in biopreservation is the trend towards 'clean label' and more 'natural' foods because of rising consumer distrust in traditional, artificial additives and preservatives (8). Bacteriophages (phages) are attracting interest for use in food preservation and food safety (9). Phages are effective at killing bacteria, are specific to their host, and, as a result of their ubiquity in nature, are commonly encountered and ingested, making them attractive for use in food safety. The anti-*Listeria* phage Listex P100 contains a single phage, P100, and was granted Generally Regarded as Safe (GRAS) status by the American Food and Drug Agency (FDA) in 2006 (10). Bacteriocins are another alternative to traditional additives and preservatives for food safety (11). Bacteriocins are ribosomally synthesised antimicrobial peptides with a broad or narrow spectrum of action produced by bacteria. Many bacteriocins are produced by lactic acid bacteria used in food fermentations that have been granted Generally Regarded as Safe

(GRAS) status by the FDA or have been granted Qualified Presumption of Safety (QPS) by The European Food Safety Authority (EFSA) (12). Nisaplin® containing nisin has been approved as a food preservative. Here, the activity of two commercially available 'clean label' antimicrobial compounds were investigated against *L. monocytogenes*, a foodborne pathogen of great importance and increasing incidence, in coleslaw. The objectives of this study were to [1] investigate the effect of P100 alone at a high MOI (multiplicity of infection) in coleslaw, [2] find a suitable combination of concentrations of P100 and Nisaplin®, and [3] investigate the efficacy of the combination in comparison to an untreated control, P100 alone and Nisaplin® alone.

Materials and Methods

Bacteriophage Propagation and Bacteriophage Assays

Bacterial strains were grown in TSB (tryptic soy broth) and TSA (tryptic soy broth 1.5% agar w/v) at 37˚C. Plaque assays were carried out by overlaying 4 ml TSA (0.4% agar w/v) supplemented with Calcium boroglucinate (final concentration 10 mM). Plates were incubated at 30˚C for 24 h. P100 was isolated from Listex. P100 was propagated on *L. monocytogenes* strain 33116 ATCC 19117. To propagate P100, plaque assays were carried out. After 24 h incubation, 5 ml SM buffer (50 mM Tris-HCl; 100 mM NaCl; 8.5 mM MgSO4; pH 7.5) and 2% (v/v) chloroform were added to plates with confluent plaques and incubated at room temperature for 10 min. Phage suspension was removed from the plates, centrifuged at 4700 X g for 10 min at 4˚C in a swing-bucket centrifuge, and filtered twice through a 0.45 µm pore diameter filter. Chloroform was removed using a VivaSpin® 6 column.

Host Range of P100

The host range of P100 was established by carrying out plaque assays and comparing the phage titre on the strain to be tested to the phage titre on strain 33116 ATCC 19117 (Table 1) (Titre on test strain/Titre on strain 33116 ATCC 19117).

Coleslaw Food Trial with P100

Coleslaw was purchased from a supermarket. *L. monocytogenes* ScottA was grown overnight in TSB at 37˚C, centrifuged at 5500 X g for 20 min, and the resulting pellet resuspended in PBS. Resuspended ScottA was then diluted to ~10^7 CFU/ml in PBS and 100 µl added per 10 g of coleslaw. P100 was diluted to ~10^8 PFU/ml in SM buffer. The control group was treated with 1 ml SM buffer per 10 g of coleslaw, and the test group was treated with 1 ml diluted P100 per 10 g of coleslaw. Bacteria and phage were added immediately after one another. PBS and SM buffer were added in place of ScottA and P100 as negative controls, which also served as a means for sterility monitoring throughout the experiment. Coleslaw was manually mixed for 1 min after these additions; 10 g samples of coleslaw were aliquoted and stored at 4˚C for 10 days. Then, 1 ml of liquid was pipetted off from the top of the coleslaw sample and used for spread plating of undiluted coleslaw to count low numbers of ScottA and diluting to count P100. To count P100, 500 µl of coleslaw liquid was added to 4.5 ml SM buffer, vortexed, centrifuged at 4700 X g for 10 min at 4˚C in a swing-bucket centrifuge and filtered through a 0.45 µm pore diameter filter, and spot assays carried out against ScottA. To count ScottA, 90 ml PBS was added to 10 g of coleslaw and stomached for 90 s, and then serial dilutions were prepared in PBS. Oxford agar and Oxford Listeria Selective Supplement were used to count ScottA by spread plating 100 µl of the relevant dilution. Plates were incubated at 37˚C for 24 h. To check for

resistance, colonies were streaked on Oxford agar and Oxford Listeria Selective Supplement from Day 8 and Day 10 food trial count plates and spot assays carried out in triplicate against P100 lysates.

Rate of Resistance to P100

The rate that resistance to P100 occurs was assessed using the efficiency of lysogeny protocol of Dalmasso, *et al*. (42). Plates were incubated at 30°C for 24 h. The percentage of bacteriophage insensitivity was calculated as follows: (CFU on phage seeded plates/CFU on phage-free control plates) X 100. Twenty colonies were picked from the phage seeded plate and spot assays carried out to evaluate the colonies' sensitivity to P100.

P100 and Nisaplin® Checkerboard Assay in Broth

P100 and Nisaplin® checkerboard assay was performed in TSB as per Draper, *et al*. in triplicate (43). A 2-fold serial dilution of Nisaplin® was made horizontally in sterile water (50 µl) in a 96-well microtitre plate. A 10-fold serial dilution of P100 was made vertically in SM buffer in a microtitre plate and 50 µl added vertically to the Nisaplin® dilution plate. Bacteria were grown overnight, subcultured into TSB, and allowed to grow to an OD600 nm of ~0.5. Bacteria were then diluted and 100 µl added to each microtitre well, resulting in a final concentration of 10^5 CFU/ml. The final concentrations of Nisaplin® were from 0 to 1600 µg/ml and P100 from MOI 0 to MOI 100, with the first well containing no antimicrobial agents. Wells containing media without bacteria were used to check for sterility. Growth was assessed visually after incubation for 24 h at 37˚C.

P100 and Nisaplin® Checkerboard Assay in Coleslaw Liquid

Fresh coleslaw was centrifuged at 5400 X g for 5 min at 4˚C in a swing-bucket centrifuge to separate solid and liquid parts. Then, 1 ml of coleslaw liquid was added to 24-well plates. ScottA was grown overnight in TSB at 37˚C, centrifuged at 5500 X g for 20 min, and the resulting pellet resuspended in PBS. Resuspended ScottA was then diluted to ~10^7 CFU/ml in PBS and 10 µl added per 1 ml of coleslaw. The starting concentration of ScottA in coleslaw liquid was ~10^5 CFU/ml. A 1:10 serial dilution of P100 was made horizontally in SM buffer in a microtitre plate. In a second microtitre plate, Nisaplin® dilutions were made in sterile water. Next, 10 µl of P100 and 20 µl of Nisaplin® was added per 1 ml of coleslaw. The plate was stored at 4˚C for 24 h. After 24 h, 20 µl from each sample was serially diluted in 180 µl PBS in 96-well plates, and 10 µl of coleslaw dilutions were pipetted on Oxford agar and Oxford Listeria Selective Supplement and allowed to dry. Plates were incubated at 37˚C for 24 h. The combination of P100 and Nisaplin® was evaluated using the Fractional Inhibitory Concentration (FIC) index (43). The FIC looks at the interaction of antimicrobial compounds in the inhibition of a bacterial strain. It is defined by the equation: FIC = FICX + FICY = (X/MICX) + (Y/MICY). (MICx) is the minimum inhibitory concentration of the antimicrobial alone, while (X) is the lowest level of antimicrobial X in combination with another to achieve an inhibitory effect. FIC index results are interpreted as follows: FIC $\leq$ 0.5 is synergy, 0.5 < FIC $\leq$ 0.75 is partial synergy, 0.75 < FIC $\leq$ 1.0 is additive, FIC > 1.0 is indifferent, and FIC > 4 is antagonistic.

Coleslaw Food Trial with P100 and Nisaplin® in Combination

P100 and Nisaplin® combination food trial was carried out as per the P100 food trial with some adjustments. ScottA was diluted to ~10^7 CFU/ml in PBS and 100 µl added per 10 g of coleslaw, as previously described. P100 was diluted to ~10^7 PFU/ml in SM buffer. A 500 µg/ml stock solution of Nisaplin® was prepared in sterile water, and 500 µl SM buffer and 500 µl sterile water per 10 g of coleslaw was added to the negative control samples. Then, 500 µl diluted P100 and 500 µl sterile water was added per 10 g of coleslaw to the P100 only samples; 500 µl SM buffer and 500 µl Nisaplin® stock solution per 10 g of coleslaw was added to the Nisaplin® only samples; and 500 µl diluted P100 and 500 µl Nisaplin® stock solution per 10 g of coleslaw was added to the P100 and Nisaplin® samples. Samples were stored at 4˚C for 10 days. To check for resistance, colonies were streaked on Oxford agar and Oxford Listeria Selective Supplement from Day 10 food trial count plates and spot assays carried out in triplicate against P100 lysates. To test for Nisaplin® sensitivity spot assays were carried out using 10 µl of 50 µg/ml solution of Nisaplin® on strain overlays similarly to phage spot assays.

Statistical Analysis

Bacterial and phage counts were determined by triplicate plating and all experiments were independently performed three times. Results are presented as mean values of these three experiments, and error bars in the figures indicate standard error of the mean (SEM). For CFU/g, graphs values were normalised before plotting. For coleslaw treated with P100 at MOI of 50, Student's t-test (unpaired, two-tailed) was used to determine the significance of differences between controls and phage-treated samples. For coleslaw treated with P100 and Nisaplin® in combination, one-way

ANOVA was used to determine the significance of differences between controls and treated samples.

Results

Host Range of P100

P100 is suited to use in food safety due to its broad host range (13). A range of strains of *Listeria* were tested for sensitivity to P100 (Table 1). These strains were originally isolated from a range of sources. P100, as expected, showed activity against many, but not all, pathogenic and non-pathogenic *Listeria* strains.

Coleslaw Food Trial with P100

Coleslaw was experimentally contaminated with ScottA at $7.1X10^5$ CFU/g. Over a 10-day period at 4˚C, ScottA was reduced approximately 10-fold to $7.6X10^4$ CFU/g in the untreated control (Fig. 1A). The addition of phage P100 at an MOI of 50 significantly reduced ScottA in coleslaw stored under the same conditions ($P \leq 0.001$). A significant reduction in ScottA was obvious within 2 h after phage addition ($P=0.0014$) and numbers continued to fall for the first 48 h. Specifically, within 2 h of addition of P100, the ScottA numbers were reduced from $7.1X10^5$ to $2.0X10^4$ CFU/g (Supp. Table 1). Listeria numbers remained low throughout the test period. The reductions were statistically significant at each sampling day, with an overall reduction from $7.1X10^5$ CFU/g to an undetectable level (less than 30 colonies) after 10 days (Supp. Table 1). P100 titre was tested throughout the 10-day trial (Fig. 1B). P100 titre remained high, at between $3.8X10^7$ PFU/g and $1.6X10^7$ PFU/g, during the 10-day trial. All 15 isolates were grown from individual colonies from Day 8 and all 20 colonies from Day 10, and were tested for resistance to P100 by spot assay (Supp. Table 2).

Less than 30 colonies were present on Day 8 and Day 10, so were not counted and used for statistical analysis; however, they were picked and tested for resistance to P100. No isolates were resistant to P100 after 8 or 10 days of the trial. All colonies showed similar efficiency of plaquing of P100 as naïve sensitive ScottA, which had not been included in the food trial.

The rate that resistance to P100 occurs was assessed against ScottA. The bacterial counts on phage seeded plates and on phage-free control plates were $4.7X10^2$ CFU/ml and $4.0X10^8$ CFU/ml, respectively. This gave a resistance rate of 0.0001%. Twenty colonies from the phage seeded plate were tested for sensitivity to P100 by spot assay (Supp. Table 3). Eleven of the 20 colonies were resistant to P100 and formed no plaques or zones on spot assay. Six of the 20 colonies showed reduced sensitivity to P100 (efficiency of plaquing<0.4) compared to naïve ScottA. Three of the 20 colonies were more sensitive to P100 (efficiency of plaquing>1) than naïve ScottA.

P100 and Nisaplin® Checkerboard Assays in Broth and Coleslaw

The minimum inhibitory concentration (MIC) of P100 and Nisaplin® was investigated in TSB incubated at 30˚C for 24 h. The starting concentration of ScottA in broth and coleslaw was ~10^5 CFU/ml. The MIC of P100 alone trended towards an MOI of 100. The MIC of Nisaplin® alone trended towards 400 µg/ml. No synergistic effect was seen when P100 and Nisaplin® were used in combination in checkerboard assays in broth.

To prepare for a food trial using P100 and Nisaplin® in combination against ScottA in coleslaw, a checkerboard assay was carried out in coleslaw stored at 4˚C for 24 h (Table 2). Survival was measured by counting CFU/ml of each combination. The

MIC for P100 occurred at an MOI of 10, while the MIC of Nisaplin® alone was 50 µg/ml. The combination of P100 at an MOI of 1 and Nisaplin® at a concentration of 25 µg/ml reduced the amount of each antimicrobial required to inhibit ScottA growth. The FIC of P100 at an MOI of 1 in combination with Nisaplin® at a concentration of 25 µg/ml was 0.6, which represents partial synergy against ScottA. Based on this data, the concentrations of P100 at an MOI of 2.5 and 25 µg/ml Nisaplin® were chosen for a food trial using P100 and Nisaplin® in combination.

Coleslaw Food Trial with P100 and Nisaplin® in Combination

Coleslaw was experimentally contaminated with ScottA at $9.4X10^5$ CFU/g. Over the 10-day period at 4˚C, ScottA in the untreated control was reduced to $7.4X10^4$ CFU/g, representing a reduction of more than 1 log (Fig. 2A). The combination of P100 at an MOI of 2.5 and 25 µg/ml Nisaplin® significantly reduced ($P \leq 0.001$) the ScottA burden in coleslaw stored at 4˚C over a period of 10 days. Again, as seen with the P100 alone at an MOI of 50, a reduction in ScottA was visible 2 h after the antimicrobials were added, and reductions continued until the end of the 10-day period. Within 2 h of P100 and Nisaplin® addition, ScottA numbers were reduced from $9.4X10^5$ to $1.0X10^5$ CFU/g (Supp. Table 4). ScottA numbers were reduced to $1.2X10^2$ CFU/g by day 10.

P100 alone also significantly reduced the ScottA numbers in coleslaw ($P \leq 0.001$) over the 10-day period. This reduction was visible 2 h after P100 addition ($9.4X10^5$ to $2.5X10^5$ CFU/g). There was no statistically significant difference in the levels of reduction in ScottA between P100 alone and P100 and Nisaplin® in combination ($P > 0.05$).

Nisaplin® alone did not significantly reduce ScottA numbers over the 10-day period ($P>0.05$). However, Nisaplin® did significantly reduce ScottA numbers over the first 5 days ($P<0.01$). The reduction was not significant on days 6, 8, 10, or overall ($P>0.05$). ScottA numbers were reduced ($9.4X10^5$ to $2.0X10^5$ CFU/g) within 2 h after Nisaplin® addition.

Ten colonies from each test group on Day 10 were tested for resistance to P100 and Nisaplin® by spot assay (Supp. Table 5). No colonies were found to be resistant to P100 after 10 days of the trial. All colonies showed similar efficiency of plaquing of P100 to naïve ScottA, which had not been included in the food trial. No colonies were found to be resistant to Nisaplin® after 10 days of the trial. All colonies showed similar Nisaplin® zones of inhibition to naïve ScottA, which had not been included in the food trial.

Discussion

The "Commission Regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs" outlines the acceptable levels in the EU for common pathogenic bacteria in a range of foods and at different points in their manufacturing and storage. The limit for *L. monocytogenes* is 100 CFU/g in RTE foods able to support the growth of *L. monocytogenes*, other than those intended for infants and for special medical purposes, when products are placed on the market or an absence in 25 g before the food has left the immediate control of the manufacturer. In this experiment, 100 CFU/g was chosen as a cut off as it would be unlikely for food to leave a processing plant at the starting levels of contamination used in this experiment, and this level would probably only occur after days of growth. The Food

Safety Authority of Ireland (FSAI) "Survey on verification of compliance with Commission Regulation (EC) No 2073/2005 (12NS1)" in 2014 discussed the issue of categorisation of foods in terms of which EU criteria should apply. In the FSAI survey, coleslaw fell into food category 1.2, but if it had a shorter shelf-life (<5 days), it would have fallen into food category 1.3 and, therefore, *L. monocytogenes* would have been permitted at 100 CFU/g instead of a complete absence in 25g. The average *L. monocytogenes* contamination of foods varies greatly and it has been found in foods from 10^2 to $>10^6$ CFU/g (18).

The difference in results of the checkerboard assays in broth and coleslaw could be due to temperature and pH. The concept of hurdle technology, the combination of a number of methods of preservation, is often used in food safety (19). Alone, each hurdle may not be effective, but in combination, they can reduce bacterial growth. These hurdles can include temperature, pH, salt concentration, water activity, and preservatives. The pH of coleslaw is 3.9-4.5 (20, 21), while the pH of TSB is 7.3. The TSB checkerboard was carried out at 30˚C compared to the coleslaw checkerboard, which was carried out at 4˚C. The low storage temperature and pH of coleslaw aid in reducing bacterial growth in the product. The temperature used for the experiment and the lower pH of the coleslaw may have reduced the MICs of P100 and Nisaplin® alone and in combination in coleslaw compared to TSB. Also, it has been seen that nisin has increased activity against *L. innocua* at pH 5 compared to pH 7, which could have contributed to the difference between the checkerboard assays in broth and coleslaw (22). The differences in results between food trials could be due to the effect of different foods and conditions on P100 activity. P100 had a significant effect on the growth of *L. monocytogenes* in melon and pear slices, but not on apple slices (23). The same effect was seen in juices, in that P100 had a significant effect on

the growth of *L. monocytogenes* in melon and pear juice but not in apple juice. Phage titre decreased in apple juice but stayed constant in melon and pear juice. The pH of apple slices and juice were determined to be 3.76 and 3.70, respectively, which is outside the range recommended by EFSA of pH 5.5-9.5 and an optimum of 7.7 (24). The MICs of P100 and Nisaplin® indicated by the checkerboard assay using TSB or coleslaw differed greatly. The checkerboard in broth indicated a much higher MIC for P100 and Nisaplin® against ScottA and suggests that no synergistic killing effect would occur between them, while the same analysis using coleslaw indicated MICs 10 times lower for P100 and 8 times lower for Nisaplin®, with a synergistic effect. This could have been due to the increased growth of ScottA in broth, as it was carried out at a more optimum temperature for *Listeria* growth or that the growth media itself is optimised for rapid bacterial growth.

The liquid or solid nature of food can affect the diffusion of phages (25). Phages can become bound to components of the food, and therefore unable to bind to the bacterial target. The volume of liquid present can also be an issue as, if too low a volume is present, phage will not be able to diffuse, but too great a volume can represent too large of a barrier for phage to cross. Also a low concentration of bacteria in food may mean that bacteria and phage may never meet by diffusion if phage concentration is low. It has been estimated that it would take in the order of 1000 years for 1 phage and 1 bacterium to meet within 1 ml of liquid (25). Therefore, in general, a high concentration of phage must be used to ensure a high likelihood that phage encounter the target bacterium. A lower concentration of phage may be required if bacterial numbers are high, as seen in our study.

It was not clear if phage killing over a period of 10 days was concentration-dependent. P100 at an MOI of 50 ($P \leq 0.001$) or 2.5 ($P \leq 0.001$) significantly and

similarly reduced ScottA numbers in coleslaw over a 10-day period at 4˚C, a 5.85 log reduction and a 5.97, respectively. In comparing the reduction in ScottA between Day 0 and Day 1, where the greatest killing effect occurred, there was little difference in the killing between an MOI of 50 and an MOI of 2.5 (5.85 log reduction and 5.96 log reduction, respectively). A concentration-dependent killing of *L. monocytogenes* by P100 has been observed previously. In catfish fillets, P100 at a concentration of $2X10^3$ PFU/g did not reduce *L. monocytogenes* numbers inoculated at ~4.3log10 CFU/g, but $2X10^5$ PFU/g and $2X10^7$ PFU/g both worked, with $2X10^7$ PFU/g showing even greater reduction (26). This was also seen in tuna slices, where P100 at a concentration of $5X10^8$ PFU/ml was better at reducing *L. monocytogenes* numbers inoculated at 100 CFU/g than P100 at 3˚C at a concentration of 10^5 PFU/ml, representing MOIs of 10^6 and 10^3, respectively (27). This trend was repeated in tuna slices at 3˚C inoculated with 6 log/g *L. monocytogenes* and treated with $5X10^8$ PFU/ml and 10^5 PFU/ml, representing MOIs of 100 and 0.1, respectively, but the differences were not as pronounced. In some cases, P100 did not have a significant effect on *L. monocytogenes* at the end point of the trial, but was significant at points throughout the trial.

The reduction in ScottA was time-dependent; in both cases ScottA did not reach the EU regulation for *L. monocytogenes* in RTE foods until Day 10 of the trial. It should be borne in mind that we used an initial contamination level far in excess of anything that would be likely to occur in normal food production, and so we would expect excellent control with lower initial contaminating levels. As previously mentioned, in all cases, the greatest reduction in ScottA occurred in the first day after phage addition with a 5.85 log reduction with an MOI of 50, 5.96 log reduction with an MOI of 2.5 alone, and 5.97 log reduction $9.3X10^5$ CFU/g with an MOI of 2.5 in combination with Nisaplin®. After this point, the reduction was at a similar rate to

that of negative controls. A similar rapid and dramatic reduction has previously been seen in trials using P100 in cabbage, smoked salmon, seafood, and hotdogs (28). *L. monocytogenes* decreased from $1X10^3$ CFU/g by up to 1000-fold in hot dogs, 100-fold in mixed seafood and cabbage, and 50-fold in smoked salmon between Day 0 and 1, and this level of reduction was not seen again in each food during the 6-day trial. A rapid 10-fold reduction from 10^4 CFU/g of *L. monocytogenes* after the addition of P100 in combination with *Lactobacillus sakei* was also seen in cooked ham (29). When looking at the effect of phage contact time on bacterial reduction, Soni, Nannapaneni, and Hagens found that a contact time of 30, 60, or 120 min was more effective than 15 min contact time in reducing bacterial load on catfish fillets (26). However, contact times of 30, 60, or 120 min were no more effective than one another. The titre of P100 at an initial MOI of 50 and MOI of 2.5 was stable throughout the experiment. The titre of commercially available Listex is $2X10^{11}$ PFU/ml and it can be added to food at a concentration of $1X10^9$ PFU/g, although it should be added with the estimated reduction in mind (24). P100 could have been added at a higher titre to investigate if reduction would have occurred faster and reached the recognised EU food safety cut off earlier in the shelf life of the product.

In this experiment, Nisaplin® alone did not significantly reduce ScottA numbers over the 10-day trial ($P>0.05$). A significant reduction was not expected, as Nisaplin® alone was added at sub MIC levels as established by a checkerboard assay in coleslaw and, of course, once again, we must note the extraordinarily high initial contamination levels. A combination of P100 and Nisaplin® was significantly better than Nisaplin® alone ($P\leq0.05$). Leverentz, *et al.* found Nisaplin® to be effective at reducing *L. monocytogenes* in sliced melon and apple pieces, but P100 in combination with Nisaplin® was more effective (30). Figueiredo and Almeida found that nisin and

P100 individually inhibited growth of *L. monocytogenes* in RTE sliced ham (31). A combination of P100 and nisin was significantly better at inhibiting growth of *L. monocytogenes* than nisin alone, but not P100 alone. In the trial carried out here, P100 and Nisaplin® in combination were no more effective at reducing ScottA than P100 alone ($P>0.05$).

The EFSA Panel on Biological Hazards (24) mentions the conditions of use of P100 to be between 1 and 35˚C, with an optimum temperature of 30˚C. P100 would not form plaques on *L. monocytogenes* 33116 ATCC 19117 at 37˚C, but would at 30˚C. This is consistent with the study of Tokman, *et al*. (32). The importance of the active temperature range of a phage should not be overlooked when developing phages for food safety, as phages may be required for use at refrigeration temperatures, at room temperature (in processing plants and in non-refrigerated foods), and at higher temperatures for lab assays.

A potential consequence of using phage in biopreservation and phage therapy is the possible emergence of phage resistant mutants (33). This can be addressed in a number of ways. Firstly, bacterial mutants can often lose their phage-resistant phenotype when the phage is removed (34), because the evolution of phage resistance can reduce bacterial fitness or virulence (35). Phages also have the ability to evolve past bacterial resistance mechanisms (36). Methods to limit the emergence or reduce the impact of phage-resistant bacteria include the use of phage cocktails (37), the sequential use of different phages (28), the use of phage immediately before packaging to avoid the reintroduction of contamination that may have become phage resistant, and thorough cleaning of equipment to avoid phage resistant mutants forming and being introduced to food (38). Phage-treated products should not be entered back into the production line, such as in “old-young smearing”, where mature cheeses are used

to inoculate cheese that is in production (39). Phage can also be used in combination with other compounds, such as essential oils, to avoid bacterial resistance (40).

The advantage of using a combination of antimicrobial compounds, or even a phage cocktail, can be seen from results of the rate of resistance to P100 assay and the coleslaw trial using P100 and Nisaplin® in combination. Using an initial MOI of 50 or 2.5, no colonies resistant to P100 at an MOI of 50 were found after 8 or 10 days. The efficiency of plaquing of colonies from Day 8 and Day 10 were similar, with no major increase or decrease in efficiency of plaquing occurring between Day 8 and Day 10. However, colonies from the rate of resistance to P100 assay were found to have reduced sensitivity to P100-highlighting the fact that resistance can occur to P100, albeit only rarely. *L. monocytogenes* mutants resistant to 50 µg/ml nisin were previously found at a frequency of 10^{-6} to 10^{-8} (41). Based on these prospective mutation rates to P100 and Nisaplin®, the chance of a strain developing resistance to both P100 and Nisaplin® is extremely low. Therefore, the advantage of a combination lies in that any cells resistant to one antimicrobial that is present will be inhibited by the other and vice-versa.

Conclusions

Three objectives were outlined at the start of this study. They were [1] to investigate the effect of P100 alone at a high MOI in coleslaw against a pathogen of interest, [2] to find a suitable combination of concentrations of P100 and Nisaplin® to act against the pathogen of interest in coleslaw, and [3] to investigate the efficacy of the combination in comparison to an untreated control, P100 alone and Nisaplin® alone. P100 at an MOI of 50 significantly reduced *L. monocytogenes* numbers

(P≤0.001) in heavily contaminated coleslaw over a 10-day period and reached EU safe limits for *L. monocytogenes* in RTE-foods. A checkerboard assay was carried out and a combination of P100 at an MOI of 2.5 and 25 µg/ml Nisaplin® was selected as a combination against *L. monocytogenes* in coleslaw. This combination was then used in a food trial against *L. monocytogenes*. P100 alone significantly reduced *L. monocytogenes* numbers (P≤0.001), as did P100 and Nisaplin® in combination (P≤0.001). Nisaplin® alone had no significant effect on *L. monocytogenes* numbers. P100 and Nisaplin® in combination was more effective than Nisaplin® alone, but not P100 alone. No resistance to P100 or Nisaplin® was encountered at any point.

References

1. **Jamshidi A, Zeinali T.** 2019. Significance and Characteristics of *Listeria monocytogenes* in Poultry Products. International Journal of Food Science **2019**:1-7.

2. **Hernandez-Milian A, Payeras-Cifre A.** 2014. What is new in listeriosis? BioMed research international **2014:**358051-358051.

3. **Swaminathan B, Gerner-Smidt P.** 2007. The epidemiology of human listeriosis. Microbes and Infection **9:**1236-1243.

4. **Ricci A, Allende A, Bolton D, Chemaly M, Davies R, Fernández Escámez PS, Girones R, Herman L, Koutsoumanis K, Nørrung B, Robertson L, Ru G, Sanaa M, Simmons M, Skandamis P, Snary E, Speybroeck N, Ter Kuile B, Threlfall J, Wahlström H, Takkinen J, Wagner M, Arcella D, Da Silva Felicio MT, Georgiadis M, Messens W, Lindqvist R.** 2018. *Listeria*

monocytogenes contamination of ready-to-eat foods and the risk for human health in the EU. EFSA Journal **16:**e05134.

5. **Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ.** 2014. *Listeria monocytogenes* Persistence in Food-Associated Environments: Epidemiology, Strain Characteristics, and Implications for Public Health. Journal of Food Protection **77:**150-170.

6. **Todd ECD, Notermans S.** 2011. Surveillance of listeriosis and its causative pathogen, *Listeria monocytogenes*. Food Control **22:**1484-1490.

7. **García P, Rodríguez L, Rodríguez A, Martínez B.** 2010. Food biopreservation: promising strategies using bacteriocins, bacteriophages and endolysins. Trends in Food Science & Technology **21:**373-382.

8. **Asioli D, Aschemann-Witzel J, Caputo V, Vecchio R, Annunziata A, Næs T, Varela P.** 2017. Making sense of the "clean label" trends: A review of consumer food choice behavior and discussion of industry implications. Food Research International **99:**58-71.

9. **Mahony J, McAuliffe O, Ross RP, van Sinderen D.** 2011. Bacteriophages as biocontrol agents of food pathogens. Current Opinion in Biotechnology **22:**157-163.

10. **Chan BK, Abedon ST, Loc-Carrillo C.** 2013. Phage cocktails and the future of phage therapy. Future Microbiol **8:**769-783.

11. **Cotter PD, Hill C, Ross RP.** 2005. Bacteriocins: developing innate immunity for food. Nature Reviews Microbiology **3:**777.

12. **Silva CCG, Silva SPM, Ribeiro SC.** 2018. Application of Bacteriocins and Protective Cultures in Dairy Food Preservation. Frontiers in microbiology **9:**594-594.

13. **Carlton RM, Noordman WH, Biswas B, de Meester ED, Loessner MJ.** 2005. Bacteriophage P100 for control of *Listeria monocytogenes* in foods: Genome sequence, bioinformatic analyses, oral toxicity study, and application. Regulatory Toxicology and Pharmacology **43:**301-312.

14. **Jaradat ZW, Bhunia AK.** 2003. Adhesion, invasion, and translocation characteristics of *Listeria monocytogenes* serotypes in Caco-2 cell and mouse models. Applied and environmental microbiology **69:**3640-3645.

15. **Clayton EM, Hill C, Cotter PD, Ross RP.** 2011. Real-time PCR assay to differentiate Listeriolysin S-positive and -negative strains of *Listeria monocytogenes*. Applied and environmental microbiology **77:**163-171.

16. **Fox EM, Leonard N, Jordan K.** 2011. Physiological and transcriptional characterization of persistent and nonpersistent *Listeria monocytogenes* isolates. Applied and environmental microbiology **77:**6559-6569.

17. **Murray EGD, Webb RA, Swann MBR.** 1926. A disease of rabbits characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillus *Bacterium monocytogenes* (n.sp.). The Journal of Pathology and Bacteriology **29:**407-439.

18. **Lianou A, Sofos JN.** 2007. A review of the incidence and transmission of *Listeria monocytogenes* in ready-to-eat products in retail and food service environments. J Food Prot **70:**2172-2198.

19. **Leistner L, Gorris LGM.** 1995. Food preservation by hurdle technology. Trends in Food Science & Technology **6:**41-46.

20. **Wu FM, Beuchat LR, Doyle MP, Garrett V, Wells JG, Swaminathan B.** 2002. Fate of *Escherichia coli* O157:H7 in coleslaw during storage. J Food Prot **65:**845-847.

21. **George AE, Levett PN.** 1990. Effect of temperature and pH on survival of *Listeria monocytogenes* in coleslaw. Int J Food Microbiol **11:**345-349.

22. **Gänzle MG, Weber S, Hammes WP.** 1999. Effect of ecological factors on the inhibitory spectrum and activity of bacteriocins. International Journal of Food Microbiology **46:**207-217.

23. **Oliveira M, Viñas I, Colàs P, Anguera M, Usall J, Abadias M.** 2014. Effectiveness of a bacteriophage in reducing *Listeria monocytogenes* on fresh-cut fruits and fruit juices. Food Microbiology **38:**137-142.

24. **EFSA Panel on Biological Hazards.** 2016. Evaluation of the safety and efficacy of Listex™ P100 for reduction of pathogens on different ready-to-eat (RTE) food products. EFSA Journal **14:**e04565.

25. **Hagens S, Loessner MJ.** 2010. Bacteriophage for Biocontrol of Foodborne Pathogens: Calculations and Considerations. Current Pharmaceutical Biotechnology **11:**58-68.

26. **Soni KA, Nannapaneni R, Hagens S.** 2010. Reduction of *Listeria monocytogenes* on the Surface of Fresh Channel Catfish Fillets by bacteriophage Listex P100. Foodborne Pathogens and Disease **VOL. 7, NO. 4**

27. **Migueis S, Saraiva C, Esteves A.** 2017. Efficacy of LISTEX P100 at Different Concentrations for Reduction of *Listeria monocytogenes* Inoculated in Sashimi. J Food Prot **80:**2094-2098.

28. **Guenther S, Huwyler D, Richard S, Loessner MJ.** 2009. Virulent Bacteriophage for Efficient Biocontrol of *Listeria monocytogenes* in Ready-To-Eat Foods. Applied and Environmental Microbiology **75:**93-100.

29. **Holck A, Berg J.** 2009. Inhibition of *Listeria monocytogenes* in cooked ham by virulent bacteriophages and protective cultures. Appl Environ Microbiol **75:**6944-6946.

30. **Leverentz B, Conway WS, Camp MJ, Janisiewicz WJ, Abuladze T, Yang M, Saftner R, Sulakvelidze A.** 2003. Biocontrol of *Listeria monocytogenes* on Fresh-Cut Produce by Treatment with Lytic Bacteriophages and a Bacteriocin. Applied and Environmental Microbiology **69:**4519-4526.

31. **Figueiredo A, Almeida R.** 2017. Antibacterial efficacy of nisin, bacteriophage P100 and sodium lactate against *Listeria monocytogenes* in ready-to-eat sliced pork ham. Braz J Microbiol **48:**724-729.

32. **Tokman JI, Kent DJ, Wiedmann M, Denes T.** 2016. Temperature Significantly Affects the Plaquing and Adsorption Efficiencies of Listeria Phages. Frontiers in microbiology **7:**631-631.

33. **Hagens S, Loessner MJ.** 2007. Application of bacteriophages for detection and control of foodborne pathogens. Applied Microbiology and Biotechnology **76:**513-519.

34. **O'Flynn G, Ross RP, Fitzgerald GF, Coffey A.** 2004. Evaluation of a Cocktail of Three Bacteriophages for Biocontrol of *Escherichia coli* O157:H7. Applied and Environmental Microbiology **70:**3417-3424.

35. **Oechslin F.** 2018. Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy. Viruses **10:**351.

36. **Chaturongakul S, Ounjai P.** 2014. Phage-host interplay: examples from tailed phages and Gram-negative bacterial pathogens. Frontiers in Microbiology **5:**442.

37. **Abuladze T, Li M, Menetrez MY, Dean T, Senecal A, Sulakvelidze A.** 2008. Bacteriophages Reduce Experimental Contamination of Hard Surfaces, Tomato, Spinach, Broccoli, and Ground Beef by *Escherichia coli* O157:H7. Applied and Environmental Microbiology **74:**6230-6238.

38. **Brovko LY, Anany H, Griffiths MW.** 2012. Chapter Six - Bacteriophages for Detection and Control of Bacterial Pathogens in Food and Food-Processing Environment, p 241-288. *In* Henry J (ed), Advances in Food and Nutrition Research, vol 67. Academic Press.

39. **Guenther S, Loessner MJ.** 2011. Bacteriophage biocontrol of *Listeria monocytogenes* on soft ripened white mold and red-smear cheeses. Bacteriophage **1:**94-100.

40. **Viazis S, Akhtar M, Feirtag J, Diez-Gonzalez F.** 2011. Reduction of *Escherichia coli* O157:H7 viability on leafy green vegetables by treatment with a bacteriophage mixture and trans-cinnamaldehyde. Food Microbiology **28:**149-157.

41. **Harris LJ, Fleming HP, Klaenhammer TR.** 1991. Sensitivity and Resistance of *Listeria monocytogenes* ATCC 19115, Scott A, and UAL500 to Nisin. Journal of Food Protection **54:**836-840.

42. **Dalmasso M, de Haas E, Neve H, Strain R, Cousin FJ, Stockdale SR, Ross RP, Hill C.** 2015. Isolation of a Novel Phage with Activity against *Streptococcus mutans* Biofilms. PLOS ONE **10:**e0138651.

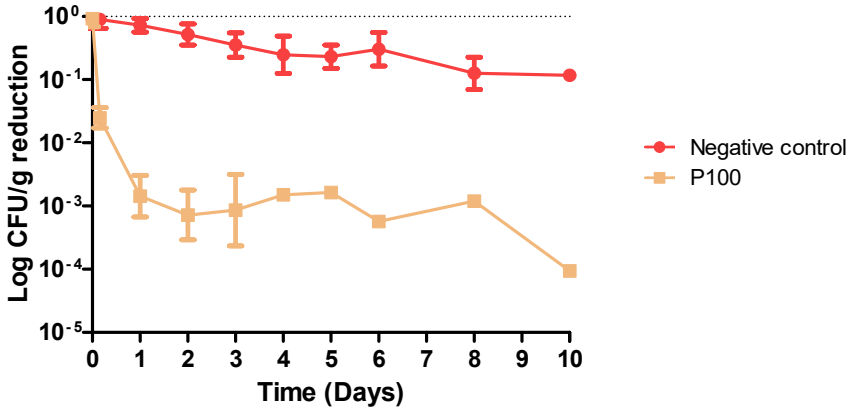
43. **Draper LA, Cotter PD, Hill C, Ross RP.** 2013. The two peptide lantibiotic lacticin 3147 acts synergistically with polymyxin to inhibit Gram negative bacteria. BMC Microbiology **13:**212.

Tables and figures

Table 1. *Listeria* strains used and their sensitivity to phage P100. Efficiency of plaquing is represented as a fraction with standard error of the mean (SEM) of three separate experiments. If no P100 plaques formed on a strain, efficiency of plaquing is represented by (-).

Strain	Equivalent Names	Origin	Serotype	Efficiency of Plaquing of P100
L. monocytogenes ATCC 35152		Guinea pig (14)	1/2a	1.22 ± 0.4
L. monocytogenes 33116	ATCC 19117	Animal (15)	4d	1 ± 0
L. monocytogenes 33120	ATCC 19118	Animal (15)	4e	0.92 ± 0.54
L. innocua DPC 3372			-	0.83 ± 0.25
L. monocytogenes 33176	20240-954	Animal (15)	1/2b	0.72 ± 0.31
L. monocytogenes 33104	F2365, JI-119, TS43	California outbreak, 1985 (15)	4b	0.70 ± 0.03
L. monocytogenes ScottA	33013	Clinical (Massachusetts outbreak, 1983) (15)	4b	0.70 ± 0.03
L. monocytogenes EGDE		Rabbit (16)	1/2a	0.61 ± 0.2
L. monocytogenes 33007	RM2218	Food (15)	4b	0.59 ± 0.22
L. monocytogenes 33186	20674-01	Animal (15)	1/2b	0.49 ± 0.25
L. monocytogenes ATCC 15313		Rabbit (17)	1/2c	0.29 ± 0.02
L. monocytogenes ATCC 19112		Human CSF (14)	1/2c	0.24 ± 0.06
L. monocytogenes 6179		Cheese; production environment (16)	1/2a	0.002 ± 0.001
L. grayi CD671	ATCC 25400	Corn stalks (15)	-	-
L. monocytogenes 33028	OB001102	Food (15)	1/2b	-
L. innocua FA2039			-	-

A.



B.

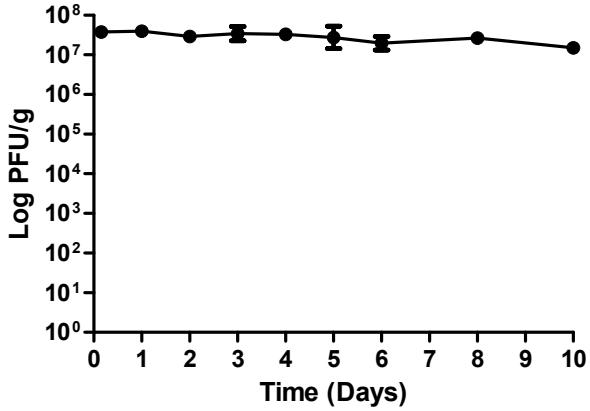


Figure 1. A. Effect of P100 on ScottA reduction in coleslaw stored at 4˚C over a 10-day period. Error bars represent SEM. Coleslaw was spiked with bacteria ($7.1X10^5$ CFU/g) and phage ($4.0X10^7$ PFU/g). B. P100 titre was also measured throughout the experiment.

Table 2. ScottA CFU/ml during P100 and Nisaplin® checkerboard assay in coleslaw. MIC of P100 alone and Nisaplin® alone are shown in bold. FIC values were calculated. FIC index results are interpreted as follows: FIC≤0.5 is synergy, 0.5<FIC≤0.75 is partial synergy, 0.75< FIC≤1.0 is additive, FIC>1.0 is indifferent, and FIC>4 is antagonistic. The partially synergistic combination of P100 and Nisaplin® is shown in red.

Antimicrobial concentration	P100 MOI 0	P100 MOI 0.01	P100 MOI 0.1	P100 MOI 1	P100 MOI 10
Nisaplin® 0 µg/ml	9.6×10^5	1.2×10^6	1.1×10^6	1.3×10^5	**$<1.0 \times 10^4$ MIC P100 Alone**
Nisaplin® 12.5 µg/ml	5.0×10^5	4.6×10^5	6.8×10^5	6.2×10^4	$<1.0 \times 10^4$ FIC 1.25 Indifferent
Nisaplin® 25 µg/ml	4.2×10^4	8.0×10^4	2.0×10^4	**$<1.0 \times 10^4$ FIC 0.6 Partial synergy**	$<1.0 \times 10^4$ FIC 1.5 Indifferent
Nisaplin® 50 µg/ml	**$<1.0 \times 10^4$ MIC Nisaplin® Alone**	$<1.0 \times 10^4$ FIC 1.001 Indifferent	$<1.0 \times 10^4$ FIC 1.01 Indifferent	$<1.0 \times 10^4$ FIC 1.1 Indifferent	$<1.0 \times 10^4$ FIC 2 Indifferent
Nisaplin® 100 µg/ml	$<1.0 \times 10^4$ FIC 2 Indifferent	$<1.0 \times 10^4$ FIC 2.001 Indifferent	$<1.0 \times 10^4$ FIC 2.01 Indifferent	$<1.0 \times 10^4$ FIC 2.1 Indifferent	$<1.0 \times 10^4$ FIC 3 Indifferent

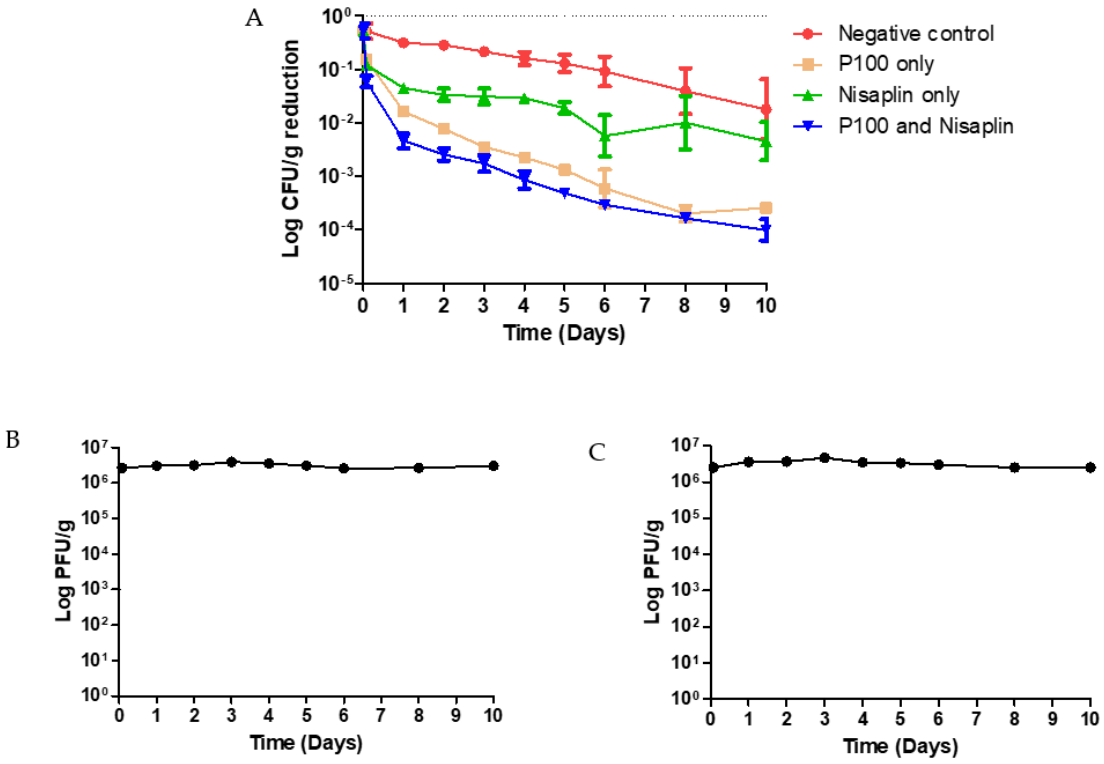


Figure 2. A. Effect of P100 and Nisaplin® in combination on ScottA reduction in coleslaw stored at 4˚C over a 10-day period. Coleslaw was spiked with bacteria ($9.4X10^5$ CFU/g) and phage (~$2.6X10^6$ PFU/g). Error bars represent SEM. P100 titre was also measured throughout the experiment of B. P100 MOI 2.5 alone and C. P100 MOI 2.5 and Nisaplin® 25 µg/ml in combination.

Supplementary Table 1. Effect of P100 on ScottA in coleslaw food trial stored at 4˚C over a 10 day period. P100 titre was also measured throughout the experiment.

Day	CFU/g	
	ScottA no P100	ScottA with phage P100
0	7.10E+05	7.10E+05
0.16	7.10E+05	2.00E+04
1	5.70E+05	1.70E+03
2	4.60E+05	1.20E+03
3	3.50E+05	8.30E+02
4	3.10E+05	3.70E+02
5	2.50E+05	4.00E+02
6	2.30E+05	1.50E+02
8	1.20E+05	2.90E+02
10	7.60E+04	0.00E+00

Day	PFU/g
	P100
0.16	3.80E+07
1	4.00E+07
2	3.00E+07
3	3.60E+07
4	3.40E+07
5	3.10E+07
6	2.10E+07
8	2.70E+07
10	1.60E+07

Supplementary Table 2. Efficiency of plaquing of 15 colonies picked from Day 8 and 15 colonies picked from Day 10 of P100 at an MOI of 50 treated coleslaw food trial to check for resistance to P100. Efficiency of plaquing is represented as a fraction with SEM of 3 separate experiments.

Colonies picked from P100 at an MOI of 50 food trial	Efficiency of plaquing
Untreated Scott A	1±0
Day 8_1	0.81±0.1
Day 8_2	0.67±0.1
Day 8_3	0.89±0.11
Day 8_4	0.71±0.04
Day 8_5	0.95±0.15
Day 8_6	0.84±0.17
Day 8_7	0.84±0.17
Day 8_8	0.89±0.11
Day 8_9	0.75±0.05
Day 8_10	0.89±0.11
Day 8_11	0.92±0.21
Day 8_12	0.83±0.26
Day 8_13	1±0.19
Day 8_14	1.03±0.17
Day 8_15	0.95±0.15
Day 10_1	0.95±0.15
Day 10_2	1±0.19
Day 10_3	0.95±0.15
Day 10_4	0.92±0.21
Day 10_5	1±0.19
Day 10_6	1±0.19
Day 10_7	1.25±0.14
Day 10_8	1.14±0.25
Day 10_9	1.14±0.25
Day 10_10	1.03±0.18
Day 10_11	1.08±0.21
Day 10_12	1.08±0.21
Day 10_13	1.14±0.25
Day 10_14	1.33±0.17
Day 10_15	1.28±0.31
Day 10_16	1.17±0.25
Day 10_17	1.22±0.28
Day 10_18	1.28±0.31
Day 10_19	1.17±0.25
Day 10_20	1.14±0.25

Supplementary Table 3. Efficiency of plaquing of colonies isolated from phage seeded plates in the rate of resistance to P100 assay. Efficiency of plaquing is represented as a fraction. If no P100 plaques formed on a colony isolated from the rate of resistance to P100 assay efficiency of plaquing is represented by (-).

Colonies isolated from efficiency of lysogeny plate	Efficiency of plaquing
Untreated ScottA	1
1	0.26
2	1.38
3	-
4	0.0003
5	0.000388
6	-
7	0.000313
8	-
9	-
10	-
11	4.88
12	-
13	-
14	-
15	0.000463
16	-
17	-
18	5.38
19	-
20	0.0005

Supplementary Table 4. Effect of P100 and Nisaplin® in combination against ScottA in coleslaw food trial stored at 4˚C over a 10 day period. P100 titre was also measured throughout the experiment.

Day	CFU/g			
	No P100 No Nisaplin®	P100 only	Nisaplin® only	P100 and Nisaplin®
0	9.40E+05	9.40E+05	9.40E+05	9.40E+05
0.16	9.40E+05	2.50E+05	2.00E+05	1.00E+05
1	5.10E+05	2.70E+04	7.40E+04	8.20E+03
2	4.60E+05	1.20E+04	5.70E+04	4.40E+03
3	3.50E+05	5.90E+03	5.50E+04	3.10E+03
4	2.80E+05	3.70E+03	4.70E+04	1.60E+03
5	2.40E+05	2.20E+03	3.20E+04	8.00E+02
6	2.00E+05	1.50E+03	1.80E+04	4.80E+02
8	1.20E+05	3.60E+02	2.80E+04	2.60E+02
10	7.40E+04	2.10E+02	1.40E+04	1.20E+02

Day	PFU/g	
	P100 only	P100 and Nisaplin®
0.16	2.70E+06	2.50E+06
1	3.10E+06	3.70E+06
2	3.20E+06	3.70E+06
3	4.00E+06	4.70E+06
4	3.70E+06	3.70E+06
5	3.10E+06	3.50E+06
6	2.60E+06	3.10E+06
8	2.70E+06	2.50E+06
10	3.10E+06	2.60E+06

Supplementary Table 5. Efficiency of plaquing and Nisaplin® sensitivity of colonies picked from Day 10 of combination food trail of Nisaplin® alone, P100 alone and P100 and Nisaplin® in combination. Efficiency of plaquing is represented as a percentage with SEM of 3 separate experiments.

Colonies picked from Day 10 of combination food trial	Efficiency of plaquing	Nisaplin® zone of inhibition diameter (mm)
Untreated ScottA	1±0	10.06
P100 alone_1	0.90±0.05	10.04
P100 alone_2	0.89±0.06	10.15
P100 alone_3	1±0	9.9
P100 alone_4	0.93±0.07	10.56
P100 alone_5	0.98±0.1	9.47
P100 alone_6	0.98±0.1	10.08
P100 alone_7	0.93±0.07	9.75
P100 alone_8	1±0	9.38
P100 alone_9	1±0	10.15
P100 alone_10	1±0	9.86
Nisaplin® alone_1	1±0	10.45
Nisaplin® alone_2	1±0	10.49
Nisaplin® alone_3	0.95±0.05	10.45
Nisaplin® alone_4	1±0	10.4
Nisaplin® alone_5	0.78±0.06	10.52
Nisaplin® alone_6	0.93±0.07	9.71
Nisaplin® alone_7	0.94±0.06	10.37
Nisaplin® alone_8	1±0	10.39
Nisaplin® alone_9	0.93±0.07	9.77
Nisaplin® alone_10	1.05±0.05	10.07
P100 and Nisaplin®_1	0.95±0.05	9.55
P100 and Nisaplin®_2	0.83±0.02	10.09
P100 and Nisaplin®_3	0.95±0.05	10.09
P100 and Nisaplin®_4	0.85±0.08	10.27
P100 and Nisaplin®_5	0.90±0.05	9.57
P100 and Nisaplin®_6	0.95±0.05	9.67
P100 and Nisaplin®_7	0.95±0.05	9.7
P100 and Nisaplin®_8	0.90±0.05	9.19
P100 and Nisaplin®_9	0.89±0.06	9.79
P100 and Nisaplin®_10	0.95±0.05	9.26

Chapter 3

The effect of hormonal contraceptives on the gut phageome over the menstrual cycle; a pilot study

Bioinformatic analysis was performed by Dr Aonghus Lavelle and Dr Stephen R Stockdale. 16S preparation and sequencing was performed by Amy Murphy.

Abstract

Metagenomics has been used to study the microbiota in both healthy and diseased humans across a number of variables in terms of diet, age and medication. Hormonal contraception is one of the most widely used group of medications but we are not aware of any studies investigating their effect on the microbiome, and particularly the viral component, the virome. In this pilot study women using hormonal contraception showed reduced virome diversity compared to women not using hormonal contraception. No differences were seen in 16S bacteriome data. Virome differences were more pronounced when viral data was clustered. Clustering was used to overcome viral inter-individual variation and the problems associated with viral 'dark matter' (viruses with no counterparts in databases). Viromes can therefore reveal subtle differences that 16S studies may overlook. These findings suggest that hormonal contraceptive use should be taken into account when interpreting metagenomic data.

Introduction

The role of the gut microbiome in human health has recently emerged as a vibrant area of research. The gut microbiome has been demonstrated to impact on conditions such as obesity, asthma and diabetes (1-3). One component of the microbiome are bacteriophages, viruses of bacteria, which are the most common biological entities on earth. Their important role in shaping bacterial composition of the gut microbiome of humans is being increasingly recognised. Bacteriophages and their potential role within the human gut have been studied in several cohorts, such as

in twins and their mothers (4), in faecal microbiota transplantations for *Clostridium difficile* infections (5) and in Inflammatory Bowel Disease (IBD) (6).

Estrogens are steroid hormones produced in the ovaries, adrenal glands and adipose tissue (7). Estrogens can act locally or circulate to target organs. The estrobolome has been described as the bacterial genes in the human gut microbiome capable of producing products that metabolise estrogens. Many small but important links have been drawn between changes in the gut microbiota and differences in sex hormone levels and gender. The estrobolome plays an important role as a regulator of circulating and excreted estrogen levels by the conjugation and deconjugation of estrogens. Estrogens have been shown to play a role in neural development, bone density, cardiovascular health, gut epithelial barrier integrity, epithelial thickness in the reproductive tract and decreasing vaginal pH (8). Colonic transit time has been shown to be shorter in men than women but no difference was found in the colonic transit time of menopausal and premenopausal women (9).

Polycystic ovary syndrome (PCOS) is a hormonal condition affecting women of reproductive age characterised by hyperandrogenism, oligo- or anovulation and polycystic ovaries on ultrasound (10). Estrogen without the opposing effect of progesterone, due to chronic anovulation, can increase the risk of endometrial cancer development in women with PCOS. Many women with PCOS are prescribed hormonal contraception to reduce this risk. PCOS patients show a reduction in phylogenetic diversity and microbiome profile shifts in comparison to healthy controls (11). In a letrozole induced PCOS rat model administration of *Lactobacillus* or faecal microbiota transplantation improved the estrous cycles and ovarian morphologies of PCOS induced rats and significantly altered hormone levels towards that of healthy controls (12).

Hormone levels during the menstrual cycle are dynamic, with a decrease in estrogen and progesterone levels during the late luteal (pre-menstrual) phase with estrogen and progesterone levels at their lowest level during menstruation (13). Many women report gastrointestinal symptoms associated with the menstrual cycle. In a study of 156 healthy women, 73% reported at least one symptom including abdominal pain, constipation, diarrhoea, nausea, or vomiting pre-menstrually while 69% reported at least one of these symptoms during menses (14). This is an increase on the rates suggested by Moore, *et al*. whereby around one third of healthy women reported perimenstrual gastrointestinal symptoms while almost half of women suffering from functional bowel disorders reported perimenstrual gastrointestinal symptoms (15). It is also noteworthy that there is a female predominance associated with IBS, with women making up two-thirds of those with Inflammatory Bowel Syndrome (IBS) (16). Women with IBS or IBD report more and worse perimenstrual gastrointestinal symptoms than healthy women, but it is unknown whether this is a flare of disease symptoms or menstrual symptoms (16). Heitkemper, *et al*. reported that the perimenstrual gastrointestinal symptoms of IBS women taking oral contraceptives did not differ from those of IBS women not taking oral contraceptives (17).

Another time of significant hormonal change occurs during pregnancy with physiological hyperestrogenemia and hyperprogesteronemia (13). There are conflicting accounts as to whether the gut microbiome changes during pregnancy. Koren, *et al*. suggested that the gut microbiota changes during pregnancy between the first and third trimesters (18). Between the first and third trimesters a decrease in the abundance of *Faecalibacterium* occurred and in the majority of women an increase in the abundance of *Proteobacteria* was observed. These changes were associated with inflammation. This was suggested to promote metabolic changes which are beneficial

during pregnancy for energy storage and growth of the foetus. Conversely DiGiulio, *et al*. found no difference in the gut microbiota during pregnancy (19).

There is little information available comparing the gut microbiota of healthy hormonal contraceptive users and controls. Progesterone and estrogen were identified as covariates associated with microbiome composition, although stool consistency was a confounder (20). However, work has been carried out investigating the vaginal microbiota. Maintenance of the vaginal pH is important for vaginal health and is controlled by the interaction of estrogen and the vaginal microbiota (21). Estrogen production causes the deposition of glycogen in the vaginal epithelium which is then metabolised to organic acids, mostly lactic acid, by the vaginal microbiota (22). This low pH and the antimicrobial nature of the bacteria are protective against colonisation by pathogens. The vaginal microbiota can be classified into 5 groups with different groups being associated with premenopausal women, perimenopausal women and postmenopausal women. These groups are often dominated by *Lactobacillus* species. The vaginal microbiota of hormonal contraceptive users has been studied, with greater vaginal microbiota stability observed in hormonal contraceptive users (23). The vaginal microbiota of hormonal contraceptive users was more likely to be dominated by *Lactobacillus* than controls. Hormonal contraceptive users have been reported to be at a lower risk of bacterial vaginosis than controls. Oral contraceptive use has been associated with an increased risk of developing Crohn's disease, while oral contraceptive use in conjunction with smoking has been associated with an increased risk of developing ulcerative colitis (24).

In order to investigate the potential effect of hormonal contraceptives on the gut microbiome of humans, a pilot study was performed. We compared the faecal bacterial (by 16S) and viral composition of ten men and ten women, five of whom

were using hormonal contraception. A decrease in diversity of the microbiome associated with the use of hormonal contraceptives was observed in the faecal phageome but not the bacteriome, showing that the phageome can be more discriminatory than bacterial 16S data.

Materials and methods

Selection of faecal samples

All study participants provided written, informed consent. The study was approved by the Clinical Research Ethics Committee of the Cork Teaching Hospitals, University College Cork. Male faecal samples were collected at one time point. Female faecal samples were collected at 3 time points over a one month period. The hormonal contraceptive using women were age matched with non-hormonal contraceptive users. Samples were collected at approximately menstrual cycle days 1, 13, and 21.

Faecal metagenomics

Preparation of faecal viral suspensions, extraction of viral DNA and library preparation was performed as per Shkoporov, *et al*. (51). This consisted of steps including the homogenisation of faecal samples in SM buffer, separation of viruses from faecal solids, virus concentration using NaCl/PEG precipitation, chloroform extraction to remove bacterial cells and PEG, DNase and RNase treatment, Proteinase K digestion and purification of viral nucleic acid by phenol:chloroform extraction and spin columns. Before library preparation viral nucleic acid was amplified using SuperScript IV Reverse Transcriptase followed by Illustra GenomiPhi, pooled and passed through a QIAGEN Blood and Tissue Purification kit. Reverse transcriptase

treatment was included to convert RNA and ssDNA viruses into dsDNA, suitable for library preparation (51). GenomiPhi was necessary to obtain sufficient DNA for library preparation due to low concentrations of extracted DNA. However, reverse transcriptase treatment can amplify bacterial rRNA contamination and GenomiPhi can preferentially amplify short single-stranded DNA such as members of the *Microviridae*. DNA concentrations were equalised and library was prepared using Nextera XT library preparation (Illumina, San Diego, CA, USA) as described by the manufacturer. Metagenomic sequencing was performed using the Illumina MiSeq platform.

16S rRNA extraction, library preparation, and sequencing was carried out by others.

Metagenomic sequencing analysis

Quality of raw reads was visualized with FastQC v0.11.5. Adaptor removal and read trimming was carried out using Trimmomatic v0.32 (52). A quality score of 20 was applied over a sliding window of 4 and a minimum length filter of 60 to retain reads following read trimming. Host sequences were identified and removed with Kraken v0.10.5 using the human genome. Levels of bacterial and archaeal contamination were estimated with sortMeRNA v2.1 (53) by aligning forward reads to the SILVA database. Assemblies were performed individually with the metaSPAdes assembler (54). Redundant contigs, at 90% identity across 90% of their length, between samples were removed. Contigs less than 1kb were filtered out and the resulting non-redundant contigs were submitted to VIRsorter in viromes mode. BLAST search of the contigs against the RefSeq Virus database was also performed. Contigs that were predicted as viral by VIRsorter or which had a significant hit against

the RefSeq Virus database were included for further analysis. The final database also included the remaining contigs which had no significant BLAST hit against the nt database and were greater than 3kb long. End-to-end alignments of sample reads were then performed against the contig database using Bowtie2 v2.1.0 and a count table was generated using SAMtools v0.1.19.

Clustering

vContact was used to form clusters with contigs grouped based on protein similarity (39). Individual genomes were analysed as representatives of viral clusters. The largest fully circular sequence from each chosen cluster was annotated. VIGA was used to annotate genomes as performed by Shkoporov, *et al*. (45). Genomes were visualised using GView and manually coloured (55). Genomes were screened for CRISPR spacers using BLASTn modified for short sequences against CRISPR spacers from Pasolli, *et al*. (56). Figure was drawn in ggplot2 in R.

Statistical analysis

Statistical analysis was performed using R v3.4.1. Alpha diversity was estimated using the Chao1 and Shannon diversity indices from vegan (package version 2.4-3). Means and standard deviation are presented. Between group comparisons were performed using the Mann-Whitney test for two-group comparisons or the Kruskal-Wallis test for more than two groups. The Hellinger transformation was used for data transformation. Bray-Curtis distances were calculated and used as input for Principal Component Analysis (PCA) using the princomp function in base R. Permutational Multivariate Analysis of Variance (PERMANOVA) and Analysis of Similarity (ANOSIM) were performed to assess all grouping variables using 99,999 permutations in vegan. Comparison between use of hormonal contraception and not

was performed with DESeq2 v1.16.1(57). Plotting was performed using ggpubr package v 0.1.5.999 with the plot and ordiplot functions.

Results

Virome assembly

Metagenomic sequencing of the viromes of ten women and ten men was carried out. Viral DNA from the faeces of ten healthy men was extracted and sequenced at one time point. Viral DNA from the faeces of the ten women was extracted and sequenced at three time points over a one month period to include the early follicular, late follicular and late luteal phases of the menstrual cycle. Five women were using hormonal contraception and five were not. A non-redundant database of 2409 putative viruses was generated and raw virome reads mapped to the assembled database (Fig. 1A). Contigs ranged in size from 1031 bp to 263205 bp (mean size 12633 bp). There were 345 circular contigs and 2064 non-circular contigs (Fig. 1B). The mean size of the circular contigs was 15447 bp (range 3006 bp to 196493 bp), although circular contigs demonstrated a bimodal distribution. The mean size of non-circular contigs was 12163 bp (range 1031 bp to 263205 bp). The number of contigs that were classified according to the various databases are as follows; 456 contigs were predicted viral by VIRsorter, 128 were identified by BLAST against the refseq viral database, 1451 contigs had greater than 3 pVOGs predicted (prokaryotic Virus Orthologous Groups) and 345 were predicted as circular genomes (Fig. 1A). There was significant overlap in these contigs and 1635 unique contigs were classified as viral by positive selection and were combined with 774 contigs identified by negative selection.

After the final putative virus database was generated vContact was used to form clusters of contigs grouped on protein similarity (Fig. 1A). Clusters were filtered to remove clusters only present in one individual or with an overall abundance of less than 0.001%. This reduced the number of clusters from 994 to 212 but retained greater than 98% of the reads, suggesting that the removed reads would contribute very little to the overall picture.

Alpha diversity

No differences were seen in alpha diversity of the bacteriome using Chao1 between any groups including men and women (Fig. 2A), the three phases of the menstrual cycle (Fig. 2A), or women using hormonal contraception and women not using hormonal contraception (Fig. 2B).

Unclustered viral contigs showed no difference in alpha diversity as measured by Chao1 between men and women or the three phases of the menstrual cycle (Fig. 2C). However, women using hormonal contraception had significantly reduced diversity compared to women not using hormonal contraception (P=0.00049) (Fig. 2D). No difference was observed between the three phases of the menstrual cycle for all women combined. However, when the phases were divided between women using hormonal contraception and women not using hormonal contraception we observed a significant reduction in diversity in women using hormonal contraception in the late luteal phase (P=0.022) (Fig. 2D).

Clustered viral data showed no difference in alpha diversity as measured by Chao1 between men and women or the three phases of the menstrual cycle (Fig. 3A). Again, women using hormonal contraception had reduced diversity compared to women not using hormonal contraception (P=0.00022) (Fig. 3B). Clustered viral data

showed no difference between the three phases of the menstrual cycle for all women combined but did show a significant reduction in women using hormonal contraception during the late follicular (P=0.037) and late luteal phases (P=0.022) (Fig. 3B).

No differences were seen in alpha diversity of the bacteriome using Shannon between men and women or the three phases of the menstrual cycle (Fig. 4A). Unlike Chao1, Shannon showed reduced diversity in women using hormonal contraception (P=0.02) compared to women not using hormonal contraception (Fig. 4B). No difference was observed when the phases were split into women using hormonal contraception and women not using hormonal contraception (Fig. 4B).

Unclustered viral contigs showed no difference in alpha diversity as measured by Shannon between men and women or the three phases of the menstrual cycle (Fig. 4C). Women using hormonal contraception had reduced diversity that was approaching statistical significance (P=0.056) compared to women not using hormonal contraception (Fig. 4D). When the phases were split into women using hormonal contraception and women not using hormonal contraception there was a no reduction in diversity in women using hormonal contraception in any of the three menstrual cycle phases (Fig. 4D).

Clustered viral data showed no difference in alpha diversity as measured by Shannon between men and women or the three phases of the menstrual cycle (Fig. 4E). There was significantly reduced diversity found in women using hormonal contraception (P=0.042) compared to women not using hormonal contraception (Fig. 4F). As with unclustered viral contigs when the phases were split into women using hormonal contraception and women not using hormonal contraception there was a no

reduction in diversity in women using hormonal contraception in any of the three menstrual cycle phases (Fig. 4F).

Beta diversity

The three time points of each woman clustered together using Bray-Curtis beta diversity of bacterial and clustered viral data (Fig. 6A, B). Beta diversity did not differentiate gender, hormonal contraceptive use status or time point (Fig. 6A, B). Women not using hormonal contraception clustered together on the left hand side of the heat map while women using hormonal contraception clustered together on the right hand side (Fig. 6C). Men were interspersed equally between women not using hormonal contraception and women using hormonal contraception. Five men were located towards the left between women not using hormonal contraception, one man was located between the women using hormonal contraception and women not using hormonal contraception and four men were located to the right of the heat map between women using hormonal contraception. This could represent the similar diversity seen between men and women not using hormonal contraception and between men and women using hormonal contraception.

Differentially abundant clusters and contigs

Lefse analysis was used to test for differentially abundant viral clusters and bacterial contigs. A number of viral clusters were suggested to be differentially abundant in women using hormonal contraception and women not using hormonal contraception (Fig. 7A). A number of bacterial contigs were differentially abundant in women using hormonal contraception and women not using hormonal contraception (Fig. 7B). The differentially abundant clusters were then correlated with all bacterial OTUs (Fig. 7C).

Analysis of individual genomes as representatives of viral clusters

The largest fully circular sequence from representatives of a selection of viral clusters were annotated (Fig. 8A). Circular genomes were selected to ensure genomes were complete. Clusters representing the phage families Myovirus (cluster 87), Siphovirus (cluster 19), Podovirus (cluster 117), Inovirus (cluster 750), and Microvirus (cluster 94) were selected. An animal virus Circovirus (cluster 90) and a human virus Papillomavirus (cluster 964) were also selected. Bacterial hosts were then inferred using CRISPR spacers targeting the viral cluster (25) (Fig. 8B).

Discussion

The first publication of the metagenomic analysis dedicated to the human gut virome appeared in 2003 (26). Since then many studies have investigated changes and alterations in the gut microbiome and the virome. Virome studies have examined many areas such as differences between the luminal and mucosal gut viromes of mice (27), differences in the viromes of children born by spontaneous vaginal delivery and caesarean section (28) and compared the gut viromes of healthy controls and irritable bowel disease patients (29). Gut microbiome studies have explored the gut microbiome in a range of situations including the development of the microbiome over time in infants (30), the effect of immunosuppressants (31) and in geographically and culturally distinct areas (32).

Sex hormones have many roles in the body. Oral contraceptives, of all types, have been associated with an increase in estrogen and sex hormone binding globulin and a decrease in testosterone and Dehydroepiandrosterone sulfate (DHEAS) (33). Estrogen may modulate the mucosal immune system and maintain intestinal barrier

function while testosterone modulates immune function. The vaginal microbiota is dominated by *Lactobacillus* spp indirectly as a result of estrogen production (34). Estrogen stimulates glycogen deposition in vaginal epithelial tissue which is metabolised by human α-amylase to maltose, maltotriose and α-dextrines, which are then metabolized to lactic acid by *Lactobacillus*. In a non-obese diabetic mouse model of type 1 diabetes, the transfer of the microbiota of adult male mice to immature female mice protected female mice from the onset of type 1 diabetes (35). It can be seen that changes in the microbiota can play a role in autoimmune disease. We therefore carried out an observational pilot study on the effect of hormonal contraception on the human gut microbiome, and particularly the virome.

There are many difficulties associated with classification of viral reads due in part to the lack of widely conserved gene markers and the fact that the majority of publically available databases are mostly composed of bacterial genomes (36). A large number of viral reads often end up unclassified. Care must be taken to include as many viral reads as possible without including non-viral reads (37). Another issue associated with viral metagenomics is the inter-individual variability of people. Minot, *et al*. showed that even when a dietary intervention was introduced inter-individual variation was still greater than intra-individual variation, however viromes did become more similar than before the intervention (38). Beta diversity of the viromes within individuals over a one year period has been found to be more similar than between individuals (4). Co-twins and their mothers showed more gut bacterial similarity than unrelated individuals while the gut viromes of co-twins and their mothers showed no more similarity than unrelated individuals. This was also apparent in the data presented here where the beta diversity of the three time points of each woman were closely grouped (Fig. 6B). Viral clustering can be used to overcome the issues of inter-

individual variability and viral dark matter (39, 40). Clustering can also be helpful to infer details about badly annotated cluster members from other more annotated members. In our study prospective hosts were suggested by CRISPR analysis (Fig. 8B) which could be helpful in identifying potential hosts for related cluster members. In this case clustering of data supported the results obtained with unclustered viral data but also identified more instances of reduced diversity. Clustered viral data showed reduced diversity during the late luteal and late follicular phases while unclustered viral contigs only showed reduced diversity during the late luteal phase.

No differences were seen in the 16S rRNA diversity of age-matched men and women in this study which contradicts previous studies. Mahnic and Rupnik found Shannon diversity to be higher in women than men (41). A combination of gender and age explained 2.2% of the inter-individual bacterial community variation using Bray-Curtis distances. However, Haro, *et al*. also found no significant difference in alpha diversity between men and women (42). No difference in the relative abundance of specific taxa between men and women was found at phylum level but there were differences at genus level. In another study when the confounding effect of stool consistency was taken into effect the impact of gender was greatly reduced (20). Estrogens showed anti-commensal activity against representative strains of the gut microbiota but the effect was not significant due to the small number of drugs tested (43). Perhaps in the pilot study carried out here bacterial diversity differences would have been evident at a species or strain level.

In this study 16S rRNA did not show differences between women using hormonal contraception and women not using hormonal contraception, but virome data did separate the two groups. There was a suggestion of clustering by hormonal contraceptive status in the heat map of viral clusters (Fig. 6C). The heat map also

supported the difference between women using and not using hormonal contraception. Men were not shown to be different to either group of women in the clustered heat map or using Chao1 alpha diversity (Fig. 3B). This inability of 16S data to show differences where virome data can has previously been seen in McCann, *et al*. where 16S rRNA did not show differences between caesarean section and standard vaginal delivered infants but virome data did discriminate the two cohorts (28). 16S rRNA shotgun metagenomic sequencing can miss differences between datasets due to low taxonomic resolution (44). The shorter the fragment the less accurate the taxonomic resolution; longer sequences are required for high taxonomic determination. Therefore the virome can be a more sensitive indicator of change.

Some viral clusters were suggested to be differentially abundant in women using hormonal contraception and women not using hormonal contraception. Bacterial correlations and CRISPR analysis agreed on host prediction (Fig. 7C and Fig. 8B). Cluster 14 was annotated to be crAssphage. Cluster 14 was correlated with *Alistipes* and CRISPR analysis predicted *Bacteroides fragilis* and *Parabacteroides distasonis* as hosts. *Bacteroides* was determined experimentally to be the host of crAssphage in 2018 (45). It has been proposed that crAssphage is actually a family of phages, with four subfamilies, which could have varying *Bacteroidetes* hosts which would support these results (46). Cluster 19 was identified as a Siphovirus with CRISPR analysis predicting the host as *Ruminococcus bicirculans* and correlating with *Lachnospira* which are both members of the order *Clostridiales*. It appears as if there was an overabundance of gram positive bacterial species associated with women using hormonal contraception as seen with Lefse analysis (Fig. 7B). Gram negative species appeared to be more common in women not using hormonal contraception.

This pilot study cannot determine if changes in the virome caused by the use of hormonal contraception are beneficial or detrimental to health. Virome changes can be significant in human health. Norman, *et al*. found the gut virome to be altered in IBD patients compared to household controls in a manner that was not explained by changes in the gut microbiome (6). Virome richness was reduced in IBD patients and virome changes were specific to ulcerative colitis and Crohn's disease patients. It was suggested that these changes could contribute to inflammation in the gut and the symptoms associated with the condition. In a study comparing twins discordant for malnutrition and concordant healthy twins viral contigs could be used to discriminate between concordant healthy twins and the "healthy" and malnourished twins of the discordant pair (47). Although they showed no symptoms of malnutrition the "healthy" discordant twin had features of an at-risk microbiota.

When choosing subjects and healthy controls for a microbiome study the impact of many factors are taken into account (48). Following the results of this pilot study, the question could be raised of whether or not the use of hormonal contraception or the onset or conclusion of taking hormonal contraception should be treated like age, diet, sex, or antibiotic use. Also it should perhaps be considered just how long the women have been using hormonal contraception (49). Women who have just started taking or stopped taking hormonal contraception could have an altered microbiota that is in flux or may have stopped due to negative effects. This could represent an underlying condition causing the negative effects or a microbiome such as the "healthy" discordant twin in the study previously mentioned (47). A National Health Statistics Report regarding 22,682 women between 2006 and 2010 reported that 82% of women had used the contraceptive pill at some time (50). One quarter of the women

interviewed who had taken the hormonal contraceptive pill had stopped due to side effects.

No effect of gender was observed on the diversity of 16S rRNA or viral data. This is contrary to published reports of increased bacterial diversity in women compared to men. However, women using hormonal contraception were found to have reduced viral diversity compared to women not using hormonal contraception. Men showed similar diversity to both groups of women. Similar trends were observed in unclustered viral contigs and viral clustered data with differences more pronounced in viral clustered data. Viral clustering was used to overcome issues of inter-individual variation and viral dark matter. This study highlights the importance of phages as a more sensitive tool for showing differences in data sets. The effect of hormonal contraceptive use was visible in the virome portion of the human gut microbiome when no effect was visible in the bacteriome. This study was carried out as a small, pilot study and included a limited number of subjects. It would be interesting to establish if shotgun metagenomics of the bacterial fraction of the microbiome or more subjects would make the effect of hormonal contraception visible in the bacteriome or if the virome is a more discriminatory tool for viewing change.

References

1. **Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.** 2009. A core gut microbiome in obese and lean twins. Nature **457:**480-484.

2. **Kostic Aleksandar D, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen A-M, Peet A, Tillmann V, Pöhö P, Mattila I, Lähdesmäki H, Franzosa Eric A, Vaarala O, de Goffau M, Harmsen H, Ilonen J, Virtanen Suvi M, Clish Clary B, Orešič M, Huttenhower C, Knip M, Xavier Ramnik J.** 2015. The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes. Cell Host & Microbe **17:**260-273.

3. **Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, Kuzeljevic B, Gold MJ, Britton HM, Lefebvre DL, Subbarao P, Mandhane P, Becker A, McNagny KM, Sears MR, Kollmann T, Mohn WW, Turvey SE, Brett Finlay B.** 2015. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Science Translational Medicine **7:**307ra152.

4. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI.** 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466:**334-338.

5. **Zuo T, Wong SH, Lam LYK, Lui R, Cheung K, Tang W, Ching J, Wu JC, Chan FK, Yu J, Sung JJ, Ng SC.** 2017. Bacteriophage Transfer during Fecal Microbiota Transplantation is Associated with Treatment Response in *Clostridium Difficile* Infection. Gastroenterology **67:**634-643.

6. **Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, Stappenbeck TS, McGovern DP, Keshavarzian A, Mutlu EA, Sauk J, Gevers D, Xavier RJ, Wang D,**

Parkes M, Virgin HW. 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell **160:**447-460.

7. **Plottel CS, Blaser MJ.** 2011. Microbiome and Malignancy. Cell host & microbe **10:**324-335.

8. **Baker JM, Al-Nakkash L, Herbst-Kralovetz MM.** 2017. Estrogen-gut microbiome axis: Physiological and clinical implications. Maturitas **103:**45-53.

9. **Meier R, C. B, P. DJ, B. MW, M. F, A. R, Y. T, R. B.** 1995. Influence of age, gender, hormonal status and smoking habits on colonic transit time. Neurogastroenterology & Motility **7:**235-238.

10. **Giudice LC.** 2006. Endometrium in PCOS: Implantation and predisposition to endocrine CA. Best Practice & Research Clinical Endocrinology & Metabolism **20:**235-244.

11. **Lindheim L, Bashir M, Münzker J, Trummer C, Zachhuber V, Leber B, Horvath A, Pieber TR, Gorkiewicz G, Stadlbauer V, Obermayer-Pietsch B.** 2017. Alterations in Gut Microbiome Composition and Barrier Function Are Associated with Reproductive and Metabolic Defects in Women with Polycystic Ovary Syndrome (PCOS): A Pilot Study. PLOS ONE **12:**e0168390.

12. **Guo Y, Qi Y, Yang X, Zhao L, Wen S, Liu Y, Tang L.** 2016. Association between Polycystic Ovary Syndrome and Gut Microbiota. PLoS ONE **11:**e0153196.

13. **Mulak A, Taché Y, Larauche M.** 2014. Sex hormones in the modulation of irritable bowel syndrome. World Journal of Gastroenterology : WJG **20:**2433-2448.

14. **Bernstein MT, Graff LA, Avery L, Palatnick C, Parnerowski K, Targownik LE.** 2014. Gastrointestinal symptoms before and during menses in healthy women. BMC Women's Health **14:**14-14.

15. **Moore J, Barlow D, Jewell D, Kennedy S.** 1998. Do gastrointestinal symptoms vary with the menstrual cycle? Br J Obstet Gynaecol **105:**1322-1325.

16. **Bharadwaj S, Barber MD, Graff LA, Shen B.** 2015. Symptomatology of irritable bowel syndrome and inflammatory bowel disease during the menstrual cycle. Gastroenterology Report **3:**185-193.

17. **Heitkemper MM, Cain KC, Jarrett ME, Burr RL, Hertig V, Bond EF.** 2003. Symptoms across the menstrual cycle in women with irritable bowel syndrome. Am J Gastroenterol **98:**420-430.

18. **Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, Bäckhed F, Isolauri E, Salminen S, Ley RE.** 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell **150:**470-480.

19. **DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DSA, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA.** 2015. Temporal and spatial variation of the

human microbiota during pregnancy. Proceedings of the National Academy of Sciences **112:**11060-11065.

20. **Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D'hoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J.** 2016. Population-level analysis of gut microbiome variation. Science **352:**560-564.

21. **Boskey ER, Cone RA, Whaley KJ, Moench TR.** 2001. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. Human Reproduction **16:**1809-1813.

22. **Brotman RM, Shardell MD, Gajer P, Fadrosh D, Chang K, Silver MI, Viscidi RP, Burke AE, Ravel J, Gravitt PE.** 2014. Association between the vaginal microbiota, menopause status, and signs of vulvovaginal atrophy. Menopause (New York, NY) **21:**450-458.

23. **Brotman RM, Gajer P, Holm J, Robinson CK, Ma B, Humphrys M, Tuddenham S, Ravel J, Ghanem KG.** 2016. 4: Hormonal contraception is associated with stability and lactobacillus-dominance of the vaginal microbiota in a two-year observational study. American Journal of Obstetrics and Gynecology **215:**S828-S829.

24. **Khalili H, Higuchi LM, Ananthakrishnan AN, Richter JM, Feskanich D, Fuchs CS, Chan AT.** 2013. Oral contraceptives, reproductive factors and risk of inflammatory bowel disease. Gut **62:**1153-1159.

25. **Stern A, Mick E, Tirosh I, Sagy O, Sorek R.** 2012. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res **22:**1985-1994.

26. **Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F.** 2003. Metagenomic Analyses of an Uncultured Viral Community from Human Feces. Journal of Bacteriology **185:**6220-6223.

27. **Kim MS, Bae JW.** 2015. Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice. Environ Microbiol doi:10.1111/1462-2920.13182.

28. **McCann A, Ryan FJ, Stockdale SR, Dalmasso M, Blake T, Ryan CA, Stanton C, Mills S, Ross PR, Hill C.** 2018. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. PeerJ **6:**e4694.

29. **Manrique P, Bolduc B, Walk ST, Oost J, Vos WM, Young MJ.** 2016. Healthy human gut phageome. Proc Natl Acad Sci USA **113:**10400-10405.

30. **Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, Warner BB, Tarr PI, Wang D, Holtz LR.** 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat Med **advance online publication**.

31. **De Vlaminck I, Khush Kiran K, Strehl C, Kohli B, Luikart H, Neff Norma F, Okamoto J, Snyder Thomas M, Cornfield David N, Nicolls Mark R, Weill D, Bernstein D, Valantine Hannah A, Quake Stephen R.** 2013. Temporal Response of the Human Virome to Immunosuppression and Antiviral Therapy. Cell **155:**1178-1187.

32. **Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI.** 2012. Human gut microbiome viewed across age and geography. Nature **486:**222-227.

33. **Khalili H.** 2016. Risk of Inflammatory Bowel Disease with Oral Contraceptives and Menopausal Hormone Therapy: Current Evidence and Future Directions. Drug safety **39:**193-197.

34. **Amabebe E, Anumba DOC.** 2018. The Vaginal Microenvironment: The Physiologic Role of Lactobacilli. Frontiers in Medicine **5**.

35. **Markle JGM, Frank DN, Mortin-Toth S, Robertson CE, Feazel LM, Rolle-Kampczyk U, von Bergen M, McCoy KD, Macpherson AJ, Danska JS.** 2013. Sex Differences in the Gut Microbiome Drive Hormone-Dependent Regulation of Autoimmunity. Science **339:**1084-1088.

36. **Roux S, Hallam SJ, Woyke T, Sullivan MB.** 2015. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. eLife **4:**e08490.

37. **Hesse U, van Heusden P, Kirby BM, Olonade I, van Zyl LJ, Trindade M.** 2017. Virome Assembly and Annotation: A Surprise in the Namib Desert. Frontiers in Microbiology **8:**13.

38. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.** 2011. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Research **21:**1616-1625.

39. **Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB.** 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ **5:**e3243-e3243.

40. **Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB.** 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nature Biotechnology **37:**632-639.

41. **Mahnic A, Rupnik M.** 2018. Different host factors are associated with patterns in bacterial and fungal gut microbiota in Slovenian healthy cohort. PLOS ONE **13:**e0209209.

42. **Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, Quintana-Navarro GM, Landa BB, Navas-Cortés JA, Tena-Sempere M, Clemente JC, López-Miranda J, Pérez-Jiménez F, Camargo A.** 2016. Intestinal Microbiota Is Influenced by Gender and Body Mass Index. PLOS ONE **11:**e0154090.

43. **Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, Patil KR, Bork P, Typas A.** 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature **555:**623.

44. **Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R.** 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Reviews Microbiology **12:**635.

45. **Shkoporov A, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C.** 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. Nature Communications **9:**4781.

46. **Guerin E, Shkoporov A, Stockdale SR, Clooney AG, Ryan FJ, Sutton TDS, Draper LA, Gonzalez-Tortuero E, Ross RP, Hill C.** 2018. Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. Cell Host & Microbe **24:**653-+.

47. **Reyes A, Blanton LV, Cao S, Zhao G, Manary M, Trehan I, Smith MI, Wang D, Virgin HW, Rohwer F, Gordon JI.** 2015. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. Proceedings of the National Academy of Sciences **112:**11941-11946.

48. **Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, Conrad M, Collman RG, Baldassano R, Bushman FD, Bittinger K.** 2017. Optimizing methods and dodging pitfalls in microbiome research. Microbiome **5:**52.

49. **Oinonen KA, Mazmanian D.** 2001. Effects of oral contraceptives on daily self-ratings of positive and negative affect. Journal of Psychosomatic Research **51:**647-658.

50. **Daniels K, Mosher WD.** 2013. Contraceptive methods women have ever used: United States, 1982-2010. Natl Health Stat Report**:**1-15.

51. **Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, McDonnell SA, Nolan JA, Sutton TDS, Dalmasso M, McCann A, Ross RP,**

Hill C. 2018. Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome **6:**68.

52. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:**2114-2120.

53. **Kopylova E, Noe L, Touzet H.** 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics **28:**3211-3217.

54. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA.** 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res **27:**824-834.

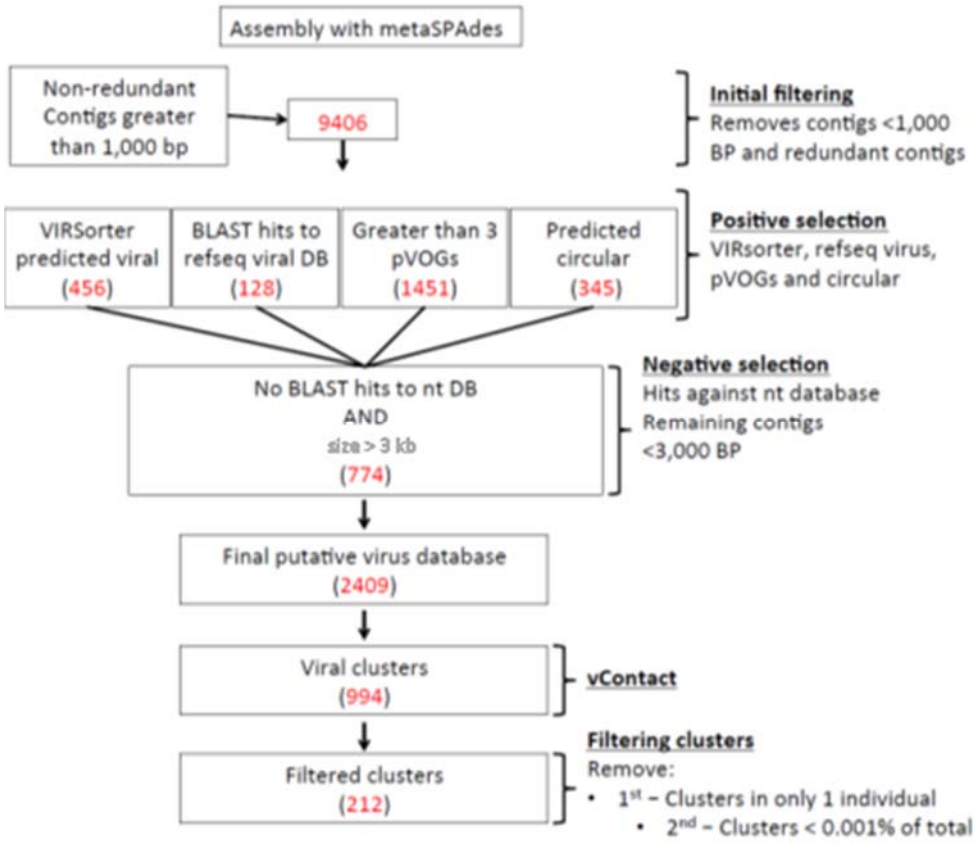
55. **Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G.** 2010. Interactive microbial genome visualization with GView. Bioinformatics **26:**3125-3126.

56. **Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N.** 2019. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell **176:**649-662.e620.

57. **Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15:**550.

Tables and figures

A.



B.

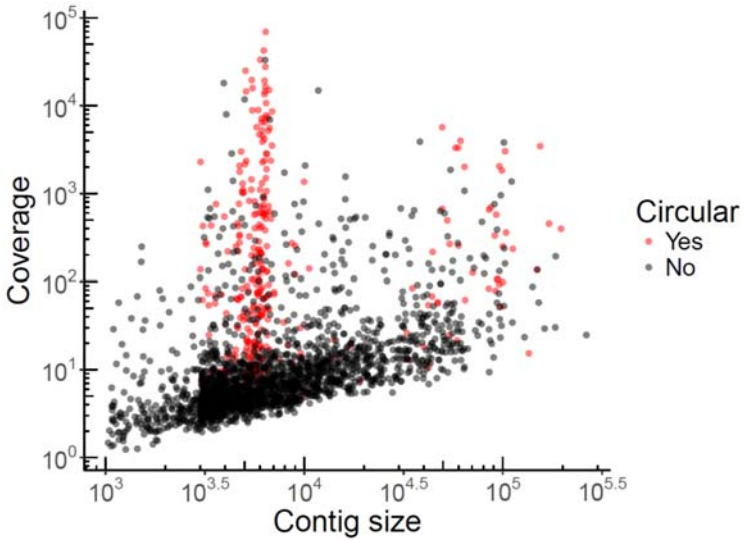


Figure 1. A. Flow chart of metagenomic analysis. B. Coverage and length of circular and non-circular predicted viral contigs. The y-axis shows the number of sequence reads, the x-axis shows contig length. Circular contigs are shown in red and non-circular contigs are shown in grey.

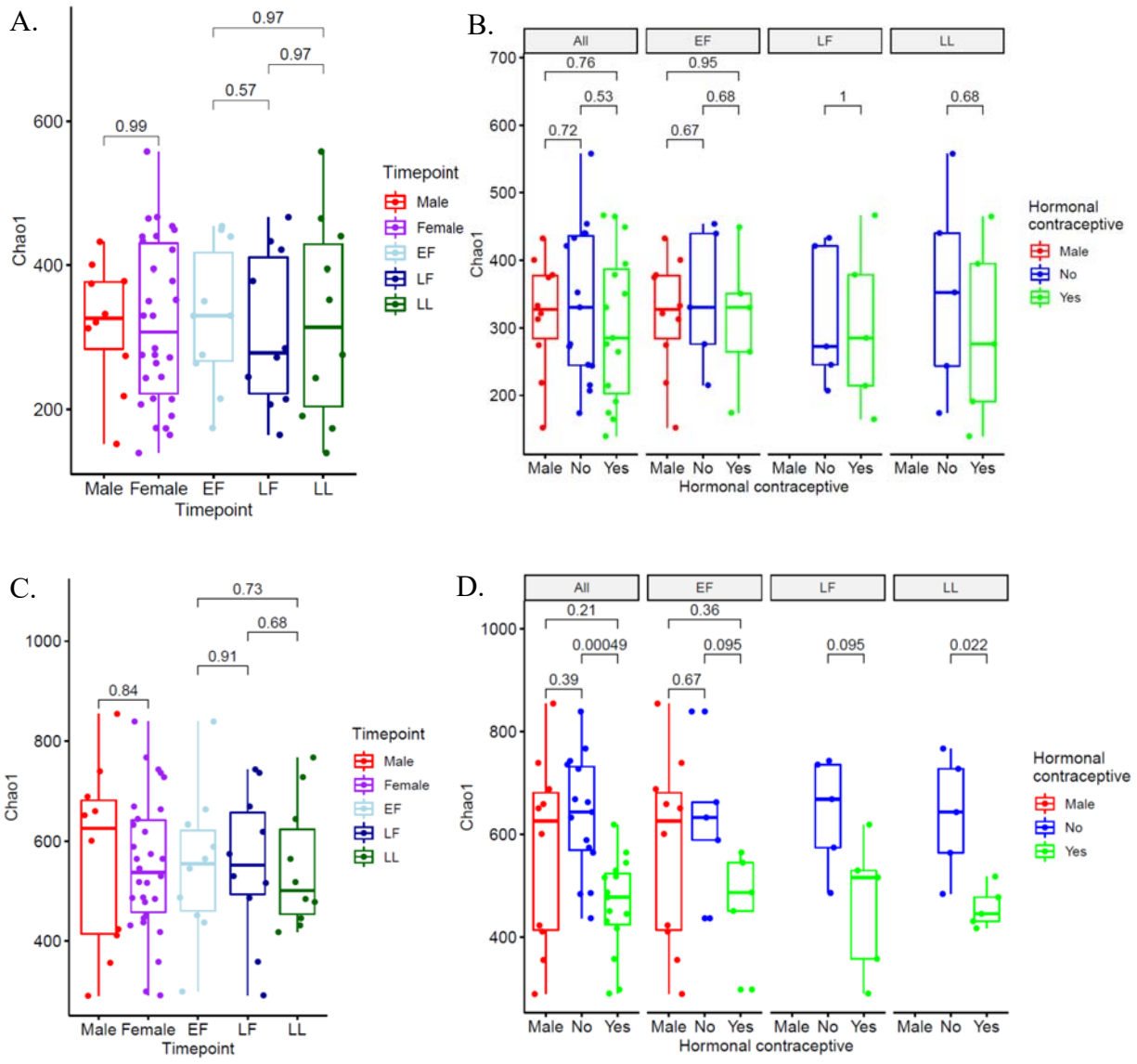


Figure 2. Chao1 Alpha diversity A. 16S rRNA comparing men, women and the three phases of the menstrual cycle. B. 16S rRNA comparing men, women using hormonal contraception, and women not using hormonal contraception over the three phases of the menstrual cycle. C. Unclustered viral contigs comparing men, women and the three phases of the menstrual cycle. D. Unclustered viral contigs comparing men, women using hormonal contraception, and women not using hormonal contraception over the three phases of the menstrual cycle.

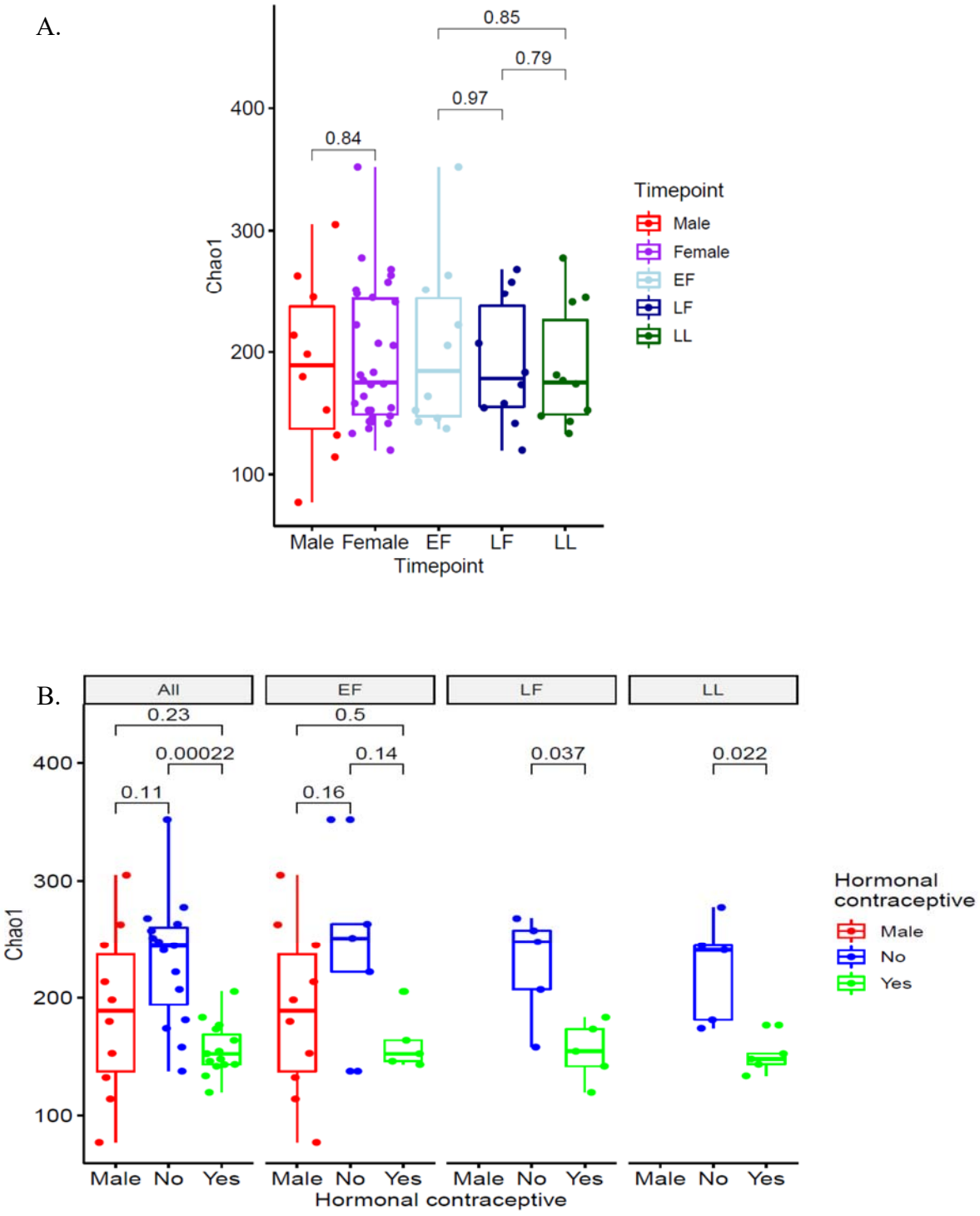
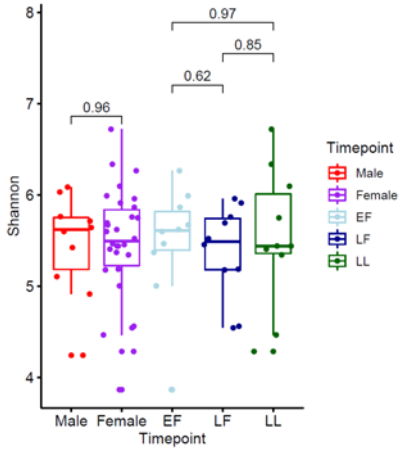
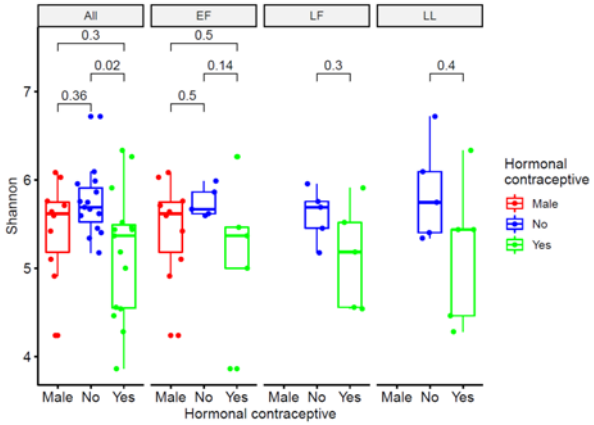


Figure 3. Chao1 Alpha diversity of clustered viral contigs comparing A. Men, women, and the three phases of the menstrual cycle B. Men, women using hormonal contraception, and women not using hormonal contraception over the three phases of the menstrual cycle.

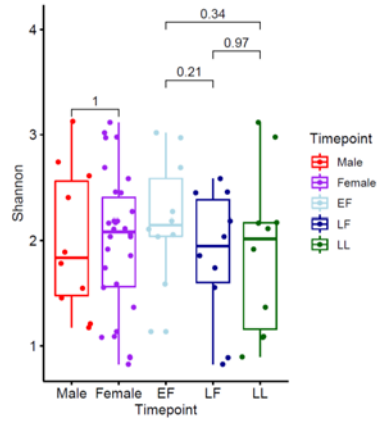
A.


0.97
0.85
0.62
0.96
Shannon
Timepoint
Male
Female
EF
LF
LL

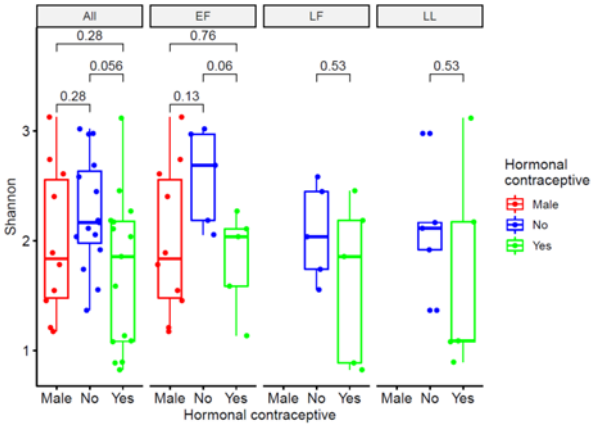
B.


All
EF
LF
LL
0.3
0.5
0.02
0.14
0.3
0.4
0.36
0.5
Shannon
Male No Yes
Hormonal contraceptive
Hormonal contraceptive
Male
No
Yes

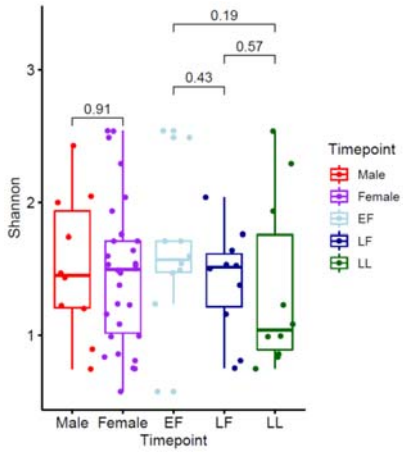
C.


0.34
0.97
0.21
1
Shannon
Timepoint
Male
Female
EF
LF
LL

D.


All
EF
LF
LL
0.28
0.76
0.056
0.06
0.53
0.53
0.28
0.13
Shannon
Male No Yes
Hormonal contraceptive
Hormonal contraceptive
Male
No
Yes

E.


0.19
0.57
0.43
0.91
Shannon
Timepoint
Male
Female
EF
LF
LL

F.

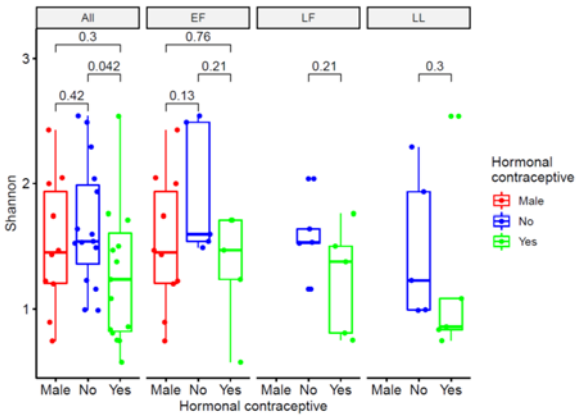

All
EF
LF
LL
0.3
0.76
0.042
0.21
0.21
0.3
0.42
0.13
Shannon
Male No Yes
Hormonal contraceptive
Hormonal contraceptive
Male
No
Yes

Figure 4. Shannon Alpha diversity A. 16S rRNA comparing men, women, and the three phases of the menstrual cycle. B. 16S rRNA comparing men, women using hormonal contraception, and women not using hormonal contraception over the three phases of the menstrual cycle. C. Unclustered viral contigs comparing men, women, and the three phases of the menstrual cycle. D. Unclustered viral contigs comparing men, women using hormonal contraception, and women not using hormonal contraception over the three phases of the menstrual cycle. E. Clustered viral contigs comparing men, women, and the three phases of the menstrual cycle. F. Clustered viral contigs comparing men, women using hormonal contraception, and women not using hormonal contraception over the three phases of the menstrual cycle.

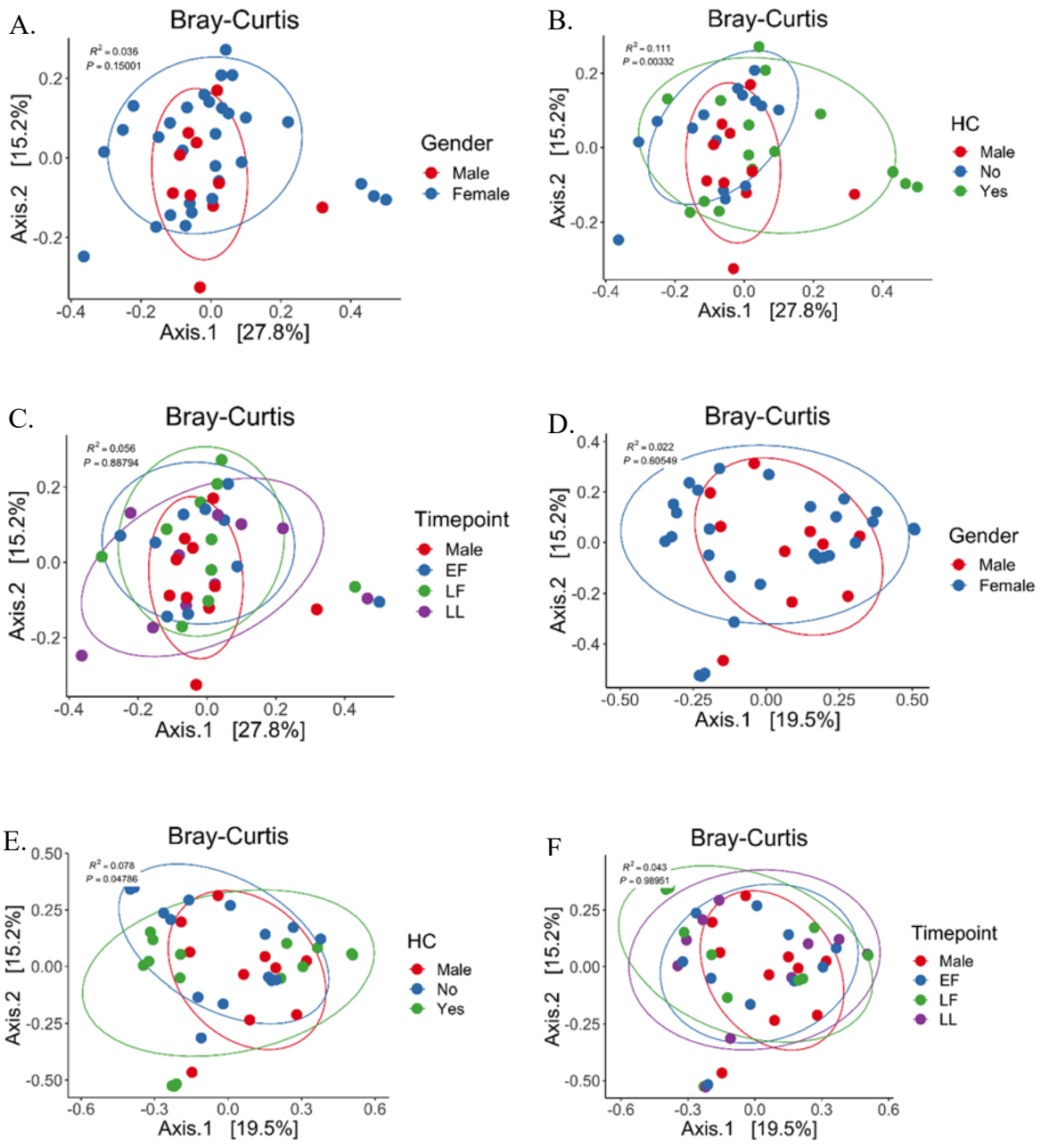


Figure 5. Beta diversity comparing A. 16S rRNA men and women. B. 16S rRNA men, women using hormonal contraception, and women not using hormonal contraception. C. 16S rRNA men and the three phases of the menstrual cycle. D. Unclustred viral contigs data men and women. E. Unclustered viral contigs men, women using hormonal contraception, and women not using hormonal contraception. F. Unclustred viral contigs men and the three phases of the menstrual cycle.

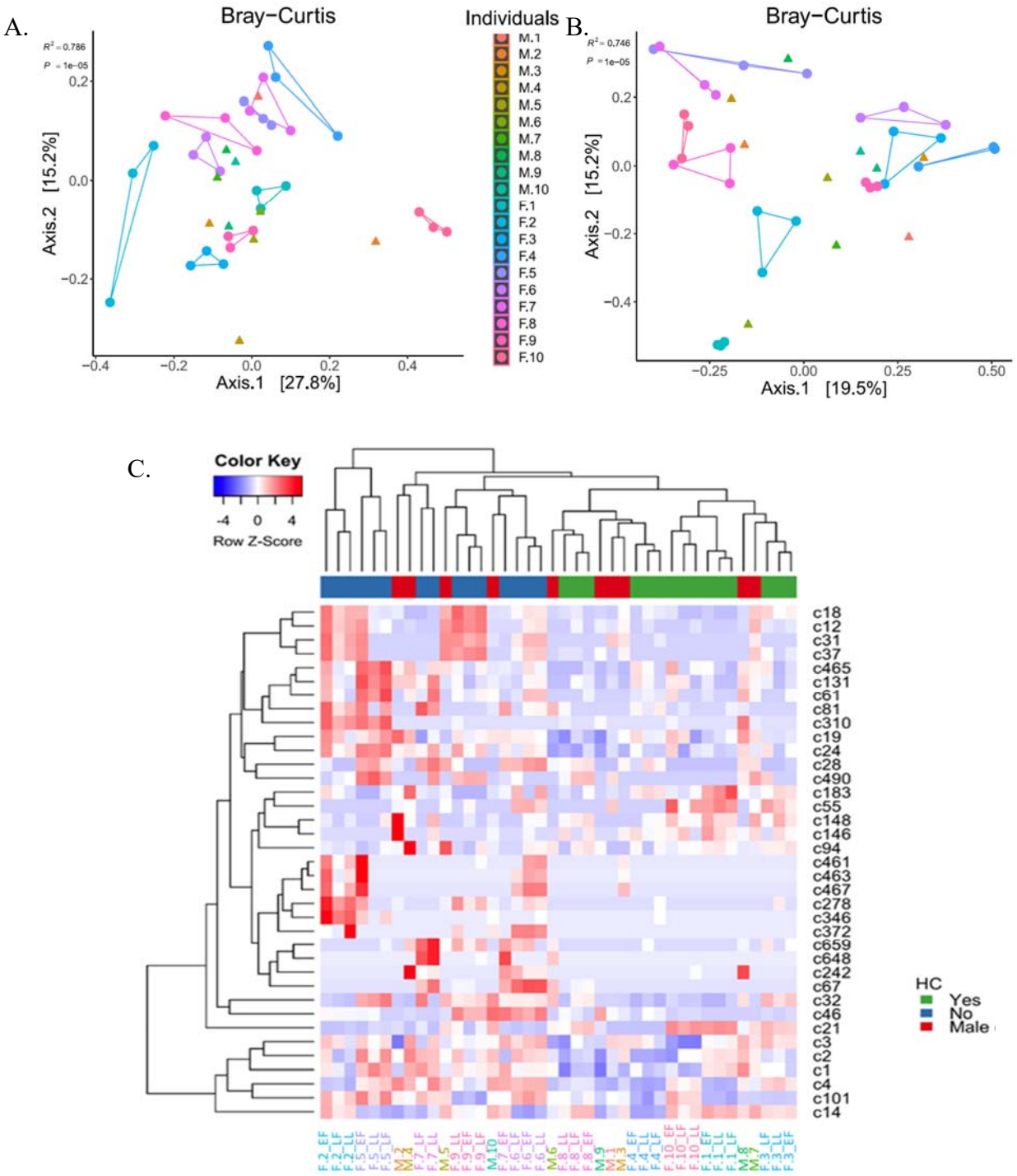
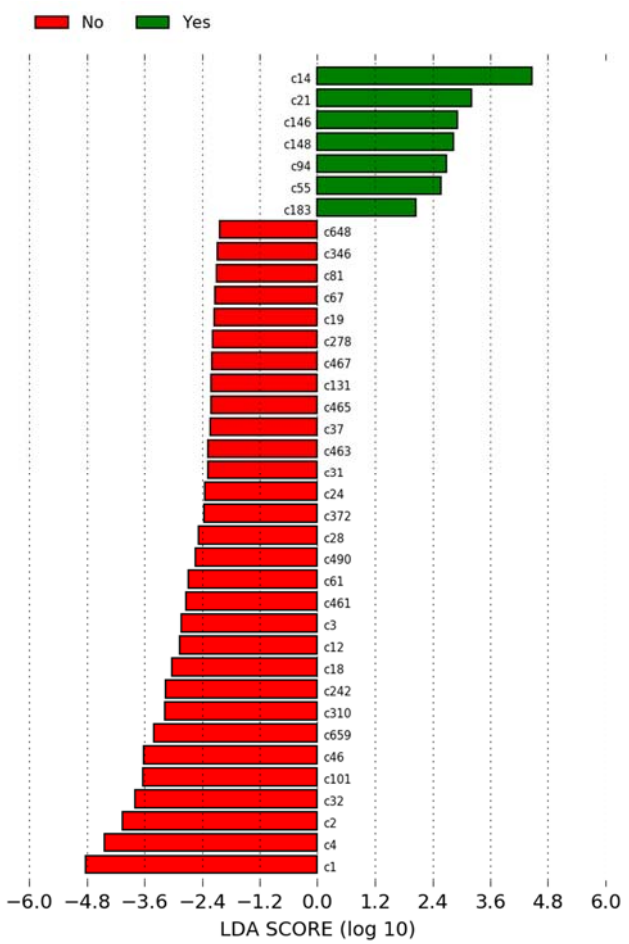
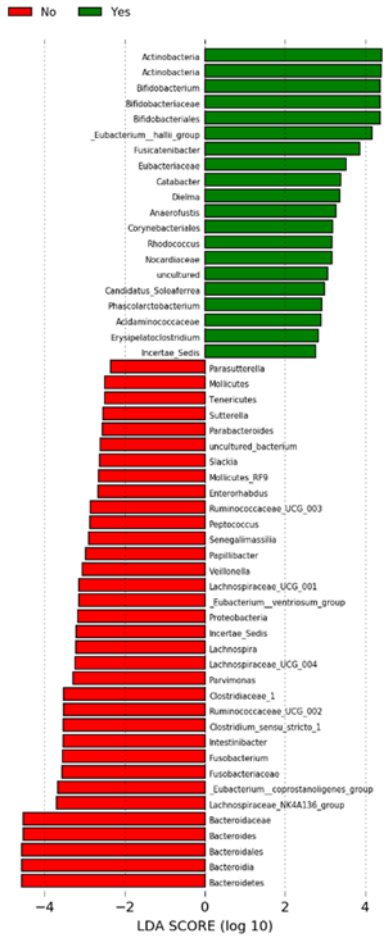


Figure 6. A. PCoA of Bray-Curtis beta diversity of individuals over three time points for women and one time point for women and men of 16S. B. PCoA of Beta diversity of individuals over three time points for women and one time point for women and men of viral clusters. Men are shown as triangles, women are shown as circles. C. Heat map of viral clusters over three time points for women and one time point for men.

A.


No
Yes
c14
c21
c146
c148
c94
c55
c183
c648
c346
c81
c67
c19
c278
c467
c131
c465
c37
c463
c31
c24
c372
c28
c490
c61
c461
c3
c12
c18
c242
c310
c659
c46
c101
c32
c2
c4
c1
−6.0 −4.8 −3.6 −2.4 −1.2 0.0 1.2 2.4 3.6 4.8 6.0
LDA SCORE (log 10)

B.


No
Yes
Actinobacteria
Actinobacteria
Bifidobacterium
Bifidobacteriaceae
Bifidobacteriales
_Eubacterium__hallii_group
Fusicatenibacter
Eubacteriaceae
Catabacter
Dielma
Anaerofustis
Corynebacteriales
Rhodococcus
Nocardiaceae
uncultured
Candidatus_Soleaferrea
Phascolarctobacterium
Acidaminococcaceae
Erysipelatoclostridium
Incertae_Sedis
Parasutterella
Mollicutes
Tenericutes
Sutterella
Parabacteroides
uncultured_bacterium
Slackia
Mollicutes_RF9
Enterorhabdus
Ruminococcaceae_UCG_003
Peptococcus
Senegalimassilia
Papillibacter
Veillonella
Lachnospiraceae_UCG_001
_Eubacterium__ventriosum_group
Proteobacteria
Incertae_Sedis
Lachnospira
Lachnospiraceae_UCG_004
Parvimonas
Clostridiaceae_1
Ruminococcaceae_UCG_002
Clostridium_sensu_stricto_1
Intestinibacter
Fusobacterium
Fusobacteriaceae
_Eubacterium__coprostanoligenes_group
Lachnospiraceae_NK4A136_group
Bacteroidaceae
Bacteroides
Bacteroidales
Bacteroidia
Bacteroidetes
−4 −2 0 2 4
LDA SCORE (log 10)

C.

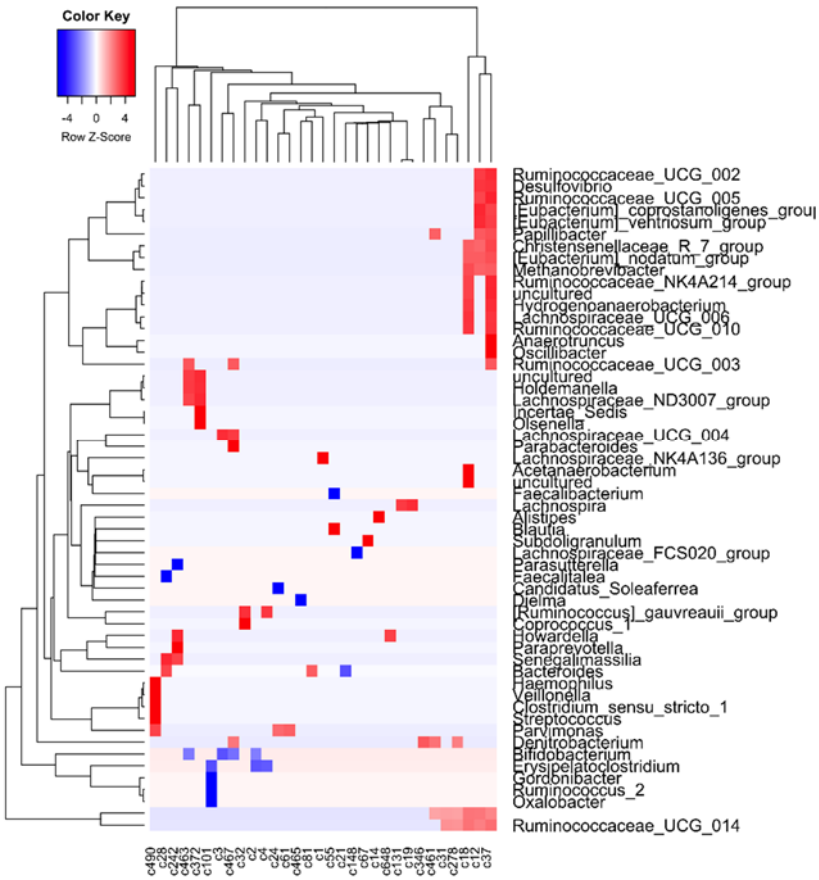
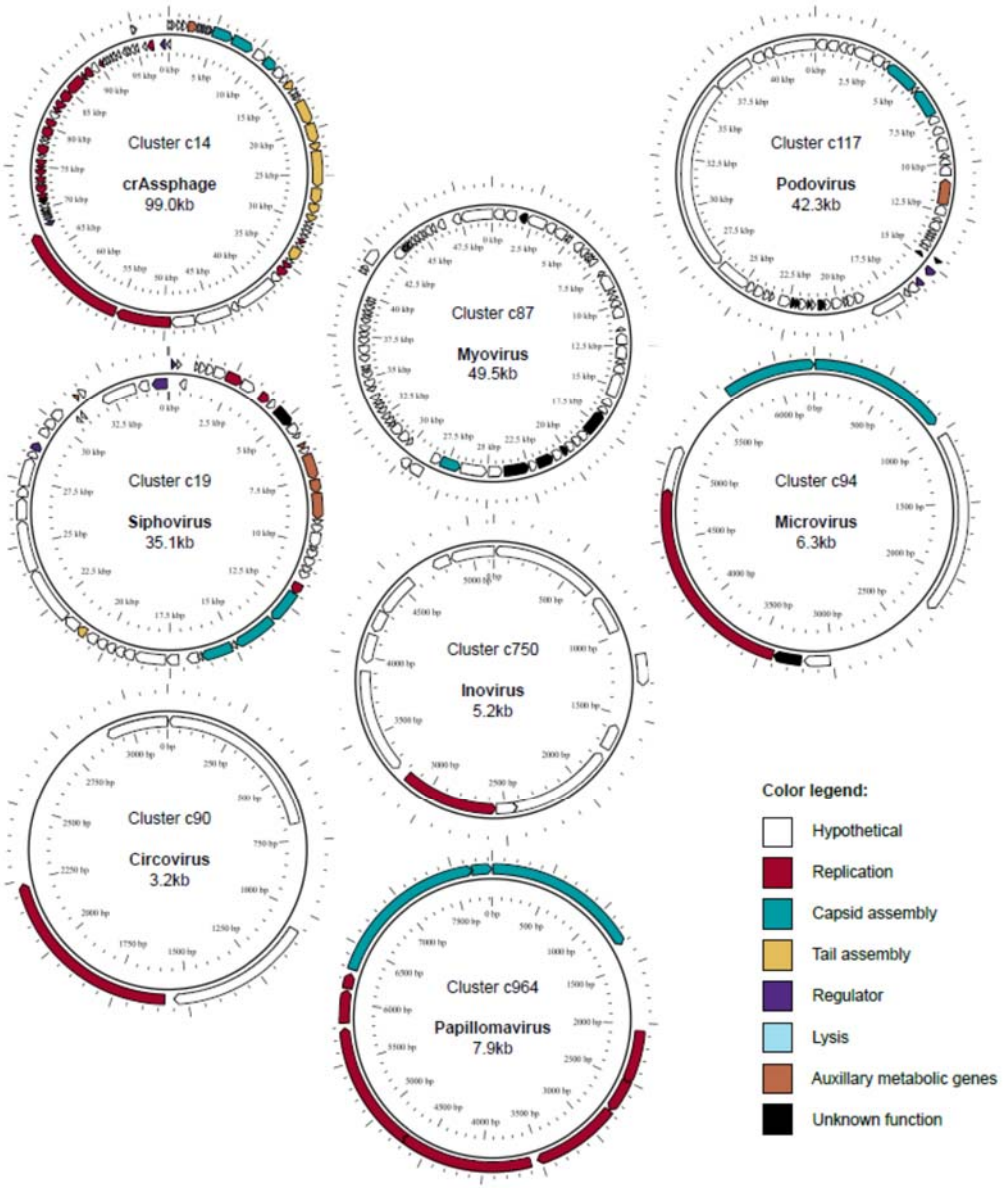

Color Key
-4 0 4
Row Z-Score
Ruminococcaceae_UCG_002
Desulfovibrio
Ruminococcaceae_UCG_005
[Eubacterium]_coprostanoligenes_group
[Eubacterium]_ventriosum_group
Papillibacter
Christensenellaceae_R_7_group
[Eubacterium]_nodatum_group
Methanobrevibacter
Ruminococcaceae_NK4A214_group
uncultured
Hydrogenoanaerobacterium
Lachnospiraceae_UCG_006
Ruminococcaceae_UCG_010
Anaerotruncus
Oscillibacter
Ruminococcaceae_UCG_003
uncultured
Holdemanella
Lachnospiraceae_ND3007_group
Incertae_Sedis
Olsenella
Lachnospiraceae_UCG_004
Parabacteroides
Lachnospiraceae_NK4A136_group
Acetanaerobacterium
uncultured
Faecalibacterium
Lachnospira
Alistipes
Blautia
Subdoligranulum
Lachnospiraceae_FCS020_group
Parasutterella
Faecalitalea
Candidatus_Soleaferrea
Dielma
[Ruminococcus]_gauvreauii_group
Coprococcus_1
Howardella
Paraprevotella
Senegalimassilia
Bacteroides
Haemophilus
Veillonella
Clostridium_sensu_stricto_1
Streptococcus
Parvimonas
Denitrobacterium
Bifidobacterium
Erysipelatoclostridium
Gordonibacter
Ruminococcus_2
Oxalobacter
Ruminococcaceae_UCG_014
c490 c28 c659 c372 c101 c3 c46 c32 c2 c4 c24 c61 c465 c81 c1 c55 c31 c148 c67 c14 c648 c131 c19 c346 c461 c37 c278 c18 c12 c37

Figure 7. A. Lefse analysis of viral clusters that explain differences between women using hormonal contraception and women not using hormonal contraception. B. Lefse analysis of bacterial groups that explain differences between women using hormonal contraception and women not using hormonal contraception. C. Heat map of viral clusters that were differentially abundant in women using hormonal contraception and women not using hormonal contraception correlated with bacterial OTUs.

A.



B.

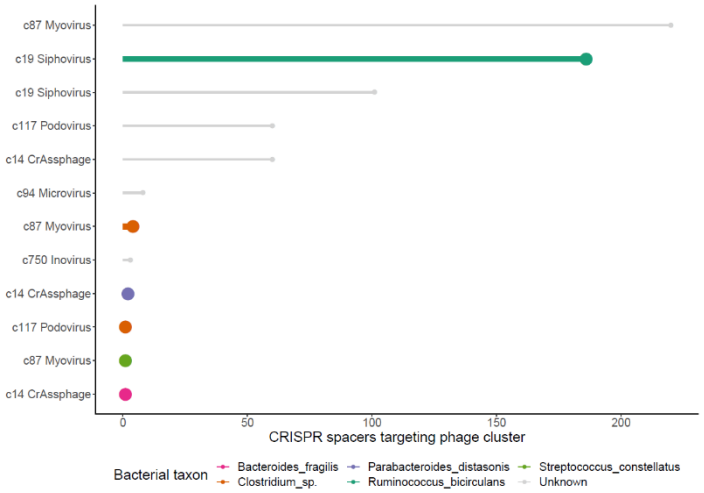


Figure 8. A. Genome maps of representatives of viral clusters. B. CRISPR spacers targeting the specific viral clusters were used to predict hosts.

Chapter 4

Analysis of single nucleotide polymorphisms in multiple strains of *Escherichia coli* bacteriophage APCEc01 isolated from different environments

Bioinformatic analysis was carried out by Dr Adam G Clooney and Dr Hugh Harris.

Abstract

Screening for new bacteriophages (phages) will be vital as increasing consumer interest in natural products and alternatives to traditional antibiotics is likely to result in an ever-increasing requirement for alternative therapies. This study involved the isolation of phages from seven different environments against a number of target strains. One novel phage was isolated, together with many previously identified phages that were re-isolated from the seven sources using 240 bacterial strains as targets, including 163 bacterial strains newly isolated in this study. Interestingly, 70 strains of phage APCEc01 were isolated against *Escherichia coli* from maize silage, grass silage, and sheep faeces. We performed single nucleotide polymorphisms (SNPs) analysis on these 70 phage isolates. Compared to the original phage genome first published in 2016 we identified 1174 different SNPs. These occurred in a range of annotated genes and hypothetical proteins. Areas of high and low SNP concentration were mapped within the genome. The number of SNPs did not correlate with the source of the phage.

Introduction

Bacteriophages (phages) are important biological tools that can potentially be used in a number of fields. Despite the increasing importance of sequencing-based, culture-independent, methods for phage identification, the isolation of phages is still as important as ever. Consumers are turning away from artificial or traditional additives and preservatives and often choosing those that they perceive as “natural” (1). Increasing numbers of products are advertised using this term. Consumers can also tend to attribute negative effects to non-natural interventions over “natural” ingredients. Foodborne illness causes significant morbidity, mortality and economic

loss every year (2). For this reason, alternative antimicrobial compounds for food safety are required, of which phages are a good option. Some phages have been investigated solely in research settings while others have been developed as commercial products such as the anti-*Listeria* product Listex, and Agriphage that targets phytopathogenic bacteria. Phages can be used at various stages of food processing to ensure food safety, such as in the reduction of contamination from livestock during production, during processing, and during storage. The rise in antibiotic resistance and the shortage of newly developed antibiotics has led to research into alternative antimicrobials for infections (3). Antimicrobial resistance is one of the greatest threats to global health. Antibiotic resistance is a naturally occurring phenomenon, but has been exacerbated by human activities such as overuse and misuse of antibiotics in human and animal health. Phages have been investigated to address this need and are seen as a viable option. Phage taxonomy is still very dependent on phage characteristics which require the culturing of phages, such as phage virion morphology for family level classification and the bacterial host for naming of phages (4).

We were interested in the natural variation between strains of phages isolated from multiple environments. Phages can encounter different selective pressures based on their environment and their mutation and evolution can be affected by this (5). The selective pressures can be due to the conditions of the environment itself or the bacteria present within it. Genetic similarity between phages can represent relatedness and shared evolution (6). Isogenic strains are often developed to use in evolution and mutation studies (7). Similar or identical phages are placed under selective pressures and allowed to evolve. The mutations and physiological effects associated with these pressures can then be tracked. We aimed to look at the occurrence and relatedness of

phage strains within and between different sources. This is one of the first studies to isolate and compare isogenic phage strains in this way.

Also of note is that sequencing and annotation can group nearly identical phages together and lose valuable information. For example, SPAdes is a commonly used assembly program in metagenomic studies. SPAdes forms a high-quality, representative, consensus sequence for almost identical sequences rather than lower quality sequences for individual strains (8). However, this can conceal variation between individual strains. These variations can be of physiological importance. In certain cases a single nucleotide polymorphism (SNP) can alter host range (9). The replication of phages is dependent on encountering a susceptible host. The less genetically related a host is to the original host the less likely the phage is to infect that host. The most commonly isolated phages have a narrow host range, limited to strains within a genus or species (10). Polyvalent phages infecting multiple species are less common. Phages are more likely to evolve under strong selective pressure, for example broad host range phages are more likely to evolve under competition for hosts (9). While having a broad host range could be seen as an evolutionary advantage, many phages persist and replicate despite having an apparently narrow host range. Bacterial hosts in turn evolve resistance to phages. Adaptation to a broader host range may have consequences. For example, mutants with increased host range have been found to have reduced infectivity towards their original hosts (11).

A selection of 86 phages isolated from human faeces, cow slurry, sheep faeces, maize silage, and grass silage, targeting 52 bacterial strains from a bank of laboratory strains and independently isolated from a healthy human faecal sample, were sequenced and analysed. All phage genomes were sequenced to determine the levels of variation across the genomes of multiple strains of the same phage.

Materials and methods

Isolation of bacterial hosts

Written consents were given according to study protocol APC055, approved by the Cork Research Ethics Committee (CREC). A healthy human faecal sample (1 g) was suspended in 5 ml of 1 X PBS and 5 ml 40% glycerol and homogenised by vortexing for 5 min. Bacterial strains were isolated by diluting the faecal suspension in PBS and spread plating 100 µl of dilutions 10^{-3}-10^{-8} on MacConkey agar (Merck), Violet Red Bile agar (Merck), Brain Heart Infusion agar (Oxoid), MRS agar (Oxoid) and LB agar (Oxoid). Agar plates were incubated at 37˚C aerobically, and also under facultative anaerobic conditions using a gas jar with an Anaerocult A (Merck). Colonies of different morphology, size and colour were isolated from various different agar plates. Strains were streaked on their respective agar plates in triplicate to purify cultures for phage screening. A bank of known strains was also selected to be used for screening (Table 2). Bacterial strains were identified by 16S rRNA sequencing analysis using BLASTn. The 16S rRNA DNA was amplified by colony PCR using primers: F8-Fw: 5'-AGAGTTTGATCMTGGCTC-3' and R1509-Rv: 5'-GNTACCTTGTTACGACTT-3'.

Isolation and purification of phages

Human faecal and farm environmental samples (2 cow slurry, 2 sheep faeces, 1 maize silage and 1 grass silage) were collected and lysates were prepared immediately. Lysates were prepared from samples using the following method. Sample (1g) was suspended in 10 ml of SM buffer (50 mM Tris-HCl; 100 mM NaCl; 8.5 mM $MgSO_4$; pH 7.5). Samples were homogenised by vortexing for 5 min, before centrifuging twice at 4,700X g for 10 min at 4˚C in a swing-bucket centrifuge to

remove large particulates and bacterial cells. Lysates were filtered twice through a 0.45 µm pore diameter filter. Bacterial strains were grown overnight in BHI at 37˚C with shaking (for aerobes) or without shaking (for facultative anaerobes). The overlay method was used for spot and plaque assays. BHI (1% agar w/v) was used as the base agar in a 100 mm X 15 mm petri dish. Plaque assays were performed by adding 400 µl of 1M $CaCl_2$ (final concentration 10mM), 100 µl of phage lysate and 100 µl of an overnight culture of bacterial host to 4 ml of soft BHI agar (0.5% agar w/v) kept at 50˚C. This mixture was poured on top of the BHI (1% agar w/v) base agar and allowed to solidify. Spot assays were performed by adding 400 µl of 1M $CaCl_2$ (final concentration 10mM) and 200 µl of an overnight culture of bacterial host to 4 ml of soft BHI agar (0.5% agar w/v) kept at 50˚C. This mixture was poured on top of the BHI (1% agar w/v) base agar and allowed to solidify. 10 µl of phage sample lysates were pipetted on the agar and allowed to dry. Plates were incubated at 37°C for 24 to 48 h. Single plaque purification by propagation of a single plaque and plaque assay was carried out 3 times. After isolation phages isolated using anaerobic conditions were grown under aerobic conditions.

Extraction of viral DNA

A 20 ml phage lysate with 4 ml of 2.5M NaCl and 50% polyethylene glycol (PEG) solution (final conc 0.4M NaCl and 8% (w/v) PEG) added was stored at 4˚C on ice overnight. Samples were centrifuged at 4700X g for 20 min at 4˚C in a swing bucket rotor. Supernatants were removed and pellets were dried for 5 min by inverting tube. Pellets were resuspended in 400 µl SM buffer (50 mM Tris-HCl; 100 mM NaCl; 8.5 mM $MgSO_4$; pH 7.5). 40 µl of 10X Nuclease Buffer (50 mM $CaCl_2$; 10 mM $MgCl_2$), was added and treated with 20 U of DNase I and 10 U of RNase I (final concentrations; Ambion) for 1 hr at 37˚C. Nucleases were inactivated at 70˚C for 10

min before samples were treated with 2 µl of freshly prepared 20 mg/µl Proteinase K for 20 min at 56˚C. Phage DNA extractions were performed using Norgen BioTek Corp Phage DNA Isolation Kit as described by the manufacturer starting at addition of Lysis Buffer B. The Elution Buffer (50 µl) was passed through the column twice to maximise DNA recovery.

Viral DNA amplification, library preparation and sequencing

Viral DNA concentrations were equalised before paired-end Nextera XT library preparation (Illumina, San Diego, CA, USA) as described by the manufacturer. Metagenomic sequencing was performed using the Illumina MiSeq (Illumina Inc., San Diego, CA, USA) by generating 300 bp paired-end read libraries following the manufacturer's instructions.

Phage genome annotation

Sequencing reads of phages were quality filtered and assembled into contigs using SPAdes (meta) (38). Redundancy was removed at 99% identity over 99% of the length. BLASTn NT database was used to compare sequences to that of known phages (39). Genomes were annotated using VIGA (https://github.com/EGTortuero/viga; (40)) and visualised using GView (41).

SARL analysis

Spot assays and plaque assays were carried out as previously described. The efficiency of plaquing was calculated by dividing the titre of each phage on the strain to be tested by the titre of that phage on strain APC 317.

The presence or absence of phage DW2 in the longitudinal data of subject 918 was tested using bowtie2 (v2.3.4.1) with default parameters.

SNP analysis of APCEc01 genomes

Reads were aligned to the reference genome APCEc01, from Dalmasso, *et al*. 2016 (12), using Bowtie2 (v2.3.4.1) with default parameters and output in SAM format. SAMtools (v1.7) was used to convert SAM files to BAM files and to sort BAM files. Bcftools (v1.7) was used to obtain coverage information (mpileup) and to call SNP variants (call--ploidy), outputting a VCF file for each sample. Vcfutils.pl (varFilter) was used to filter SNP variants by read quality and coverage, outputting filtered VCF files. Information contained in VCF files was extracted to build SNP tables for nucleotide and amino acid level. A GFF file for APCEc01 was used to match functional annotation with SNP data and also to label each SNP as either synonymous (no amino acid change) or non-synonymous (amino acid change). GView (v1.7) was used to generate a circular annotated genome that included SNP information. Barplots of synonymous and non-synonymous SNPs per gene were normalised by gene length and visualised in R (v3.6.1). Heatmaps were generated using pairwise SNP counts (nucleotide) as input and default parameters for the heatmap.2 function from the made4 package in R (v3.6.1). PCoAs were generated in R (v3.6.1) and coloured based on isolation site. Box and whisker plot of number of SNPs compared to the original APCEc01 genome from 2016 was prepared using GraphPad Prism 5.

Results

Isolation of bacterial strains and screening for phages in a human faecal sample

163 bacterial strains were isolated from a healthy human faecal sample and triple streaked to ensure pure cultures were obtained (Fig. 1 and Table 1). 37 colonies were isolated on MacConkey agar, 47 on BHI agar, 46 on MRS agar, and 24 on LB

agar. All were identified to be *Enterobacteriaceae*. As these were sequenced by 16S PCR they were not identified to strain level. 162 were identified as *Shigella/E. coli* and one as *Klebsiella aerogenes*. Seven strains were not identified. Due to the inability of 16S sequencing to distinguish between *Shigella* and *E. coli* the top three BLAST hits were included for each strain. These included *Shigella sonnei, E.fergusonii, Shigella flexneri, E.marmotae, Shigella boydii, Shigella dysenteriae.* All 240 strains, the newly isolated 163 together (Fig. 1) with a bank of 77 known strains representing various pathogens (Table 2), were screened by spot on lawn assays for phage present in faecal filtrate obtained from the same healthy human (Fig. 1). One spot was visible on *Staphylococcus aureus* APC317 and plaque purified three times. One phage (designated SARL) was isolated against *S. aureus* APC317.

Staphylococcus aureus phage DW2 analysis

SARL was sequenced and was found to resemble a previously described phage named DW2 (9) (Fig. 2). Another staphylococcal phage, CS1, was isolated in the same study as DW2 (10). SARL, DW2 and CS1 were compared (Table 3). SARL, DW2, and CS1 infected the same strains and had similar efficiencies of plaquing. SARL, DW2, and CS1 on strain *S. aureus* APC 2053 were positive on spot assay but plaques in plaque assays were hazy and unclear.

The subject from which SARL was isolated took part in a longitudinal study of the human gut virome over a twelve-month period (11). SARL was not detected in the faecal virome at any of the twelve time points, suggesting that this phage is not a constituent of the virome in this individual.

Screening for phages in environmental samples

The same 240 strains were used to screen for phages in environmental samples including maize silage, grass silage, cow faeces, and sheep faeces by spot on lawn assays (Fig. 1). 155 spots were visible against 76 bacterial strains and purified by three rounds of plaquing and picking individual plaques. Multiple plaques were picked from a single strain in some cases. 155 putative phages were isolated against 76 bacterial strains. We chose 85 phages for sequencing against 51 strains (Table 1). The majority were against *Shigella/E. coli* apart from two phages which were against *Klebsiella aerogenes*. These were selected to include a phage against as many individual bacterial strains as possible. One phage isolated was a novel phage against *Klebsiella aerogenes* which is discussed further in Chapter 5. Phages JL1, vB_EcoM_JS09, vB_Kpn_IME260, OSYSP, and Pet-CM3-4 were previously described phages targeting various *Enterobacteriaceae* including *E. coli* and *Klebsiella aerogenes* (Fig. 1). Genome maps were drawn for each phage (Fig. 3). Seventy strains of APCEc01 were isolated using *Escherichia coli* strains (Fig. 1).

Number of SNPs in APCEc01 genomes

70 strains of the phage APCEc01 were independently isolated from maize silage, grass silage, and sheep faeces. When the 70 sequenced genomes were compared to the reference genome isolated from a human faecal sample in 2016 (12), 1174 different SNPs were identified (Table 4). Any changes which were present in all 70 genomes but not the original genome was not considered a SNP, this was essentially a SNP in the original APCEc01 genome and so was discounted. There were 64 SNPs in intergenic regions and 1110 SNPs in coding regions. Of the SNPs in coding regions 286 were non-synonymous and 824 were synonymous. 88 non-synonymous SNPs

occurred in one genome only and 58 non-synonymous SNPs occurred in 69 genomes. 367 synonymous SNPs occurred in one genome only and 220 synonymous SNPs occurred in 69 genomes.

Overall the most common substitution for A was G (A→G 167/271, A→T 64/271, A→C 40/271) (Table 4). The most common substitution for C was T (C→T 234/307, C→A 51/307, C→G 22/307). The most common substitution for G was A (G→A 190/258, G→T 54/258, G→C 14/258). The most common substitution for T was C (T→C 237/338, T→A 59/338, T→G 42/338). These trends were also seen if non-synonymous substitutions and synonymous substitutions were considered separately. In the case of intergenic substitutions A to C substitution and A to T substitution were equally likely. No G to C substitutions or T to G substitutions were evident in intergenic regions but were in non-synonymous substitutions and synonymous substitutions.

SNPs did not cluster based on the source

The maximum number of SNPs compared to the original genome was 653 SNPs in A43M (Fig. 4A). After this the maximum was 527 in A1G2. The minimum number of SNPs was 387 in A1M1. The median number of SNPs was 486. Genomes isolated from maize silage, grass silage, and sheep faeces all had varying numbers of SNPs. PCoA analysis of pairwise distance between SNPs of 69 APCEc01 genomes, excluding A43M, showed no greater similarity between phage strains isolated from the same environment than strains isolated from different environments (Fig. 4B). A43M and the original APCEc01 genome were excluded as their divergence skewed the PCoA and their inclusion did not add any information as to whether the sources affected the incidence of SNPs. A43M showed the greatest variation between it and

other strains of APCEc01 isolated (Fig. 5A). When compared A43M showed between 600 and 800 SNPs between it and all of the other genomes analysed. When all other genomes were compared there were less than 300 SNPs found between any two genomes (Fig. 5B). Reduced numbers of SNPs were not seen when comparing genomes from the same source compared to different sources. No two genomes were identical. Based on the heat maps and PCoA we can establish that genomes do not cluster based on the SNPs associated with their source.

SNPs distribution across the genome

SNPs appeared to cluster rather than be regularly interspersed throughout the genome (Fig. 6). There were areas of the genome where no SNPs were present, such as in a stretch of hypothetical genes between 111000 bp and 117000 bp. A higher concentration of SNPs occurred between the start of the genome and 20000 bp. There was also a high concentration of SNPs between 148000 bp and 163000 bp. Fewer SNPs were found between these areas of SNPs from 163000 bp to 2000 bp which included a DNA helicase and a primase. Few SNPs were found in a collection of genes associated with the baseplate from 71000 bp to 78000 bp, including baseplate hub subunit, baseplate subunit, and baseplate tail tube initiator. Synonymous and non-synonymous SNPs were mostly co-occurring. Non-synonymous SNPs were present without synonymous SNPs in the same position in much fewer cases. We also noted a large number of hypothetical genes within the genome. Based on GenBank annotation APCEc01 contained 273 genes of which 55 were hypothetical proteins and 104 were phage proteins. Of the 115 genes that contained SNPs 31 were hypothetical proteins, 33 were annotated as phage proteins, and 51 had more detailed annotations.

SNPs were detected in 118 genes (Fig. 7). The number of SNPs per gene was normalised based on the length of genes. Of these 118 genes, 31 were classified as hypothetical proteins and 35 as encoding a phage protein. A number of genes only contained non-synonymous SNPs including phage holin, KEGG_ENZYME:2.4.1.18, phage baseplate hub subunit, phage major capsid protein, phage rl lysis inhibition regulator, thioredoxin, KEGG_ENZYME:2.7.7.6, phage DNA polymerase clamp subunit Gp62, seven phage proteins, and three hypothetical proteins. Molybdenum ABC transporter, phage ModA/ModB ribosyltransferase, transcriptional regulator, phage rIIA lysis inhibitor, phage nucleoid disruption protein Ndd, acridine resistance, DNA topoisomerase, KEGG_ENZYME:3.1.26.4, phage outer membrane lipoprotein Rz1, phage T4-like protein, phage RNA ligase, phage portal vertex of the head, phage baseplate wedge subunit, putative phospholipase, thioredoxin, sp spackle periplasmic protein, nine phage proteins, and 12 hypothetical proteins contained only synonymous SNPs. The other genes had a mix of synonymous and non-synonymous SNPs with the majority having a preponderance for synonymous SNPs.

The occurrence of multiple SNPs at a single site and SNPs associated with stop codons

Only in five cases were two different SNPs found at the same site (Table 5). At 18612 bp, in the phage endonuclease, both were synonymous changes which did not affect the glycine amino acid. At 18613 bp, in the phage endonuclease, A was changed to C in 66 genomes and from A to T in one genome. These were both non-synonymous changes from isoleucine to leucine and from isoleucine to phenylalanine, respectively. At 18615 bp, in the phage endonuclease, T was changed to A in 66 genomes but and T to G in one genome. The T to A change was synonymous, isoleucine, and the T to G change was non-synonymous from isoleucine to methionine. At 18633 bp, in the phage endonuclease, both were synonymous changes which did

not affect the threonine amino acid. At 34283 bp, in a hypothetical protein, T was changed to A in one genome and from T to C in one genome. These were both non-synonymous changes changing from aspartic acid to valine and glycine, respectively.

Non-synonymous substitutions introduced stop codons in four cases (Table 5). This occurred in the middle of a Phage rIIA lysis inhibitor gene (16006 bp) in two genomes. It occurred in the middle of a Putative tail fiber protein gene (27658 bp) in 67 genomes. It occurred in a hypothetical protein (53678 bp) in one genome. It also occurred in the middle of a phage protein gene (118117 bp) in 26 genomes. A non-synonymous substitution changed a stop codon to glutamic acid in a phage protein (41219 bp). There was also a substitution introduced into a stop codon but it was non-synonymous so did not have an effect. This occurred in 46 genomes and was located at the end of a phage protein gene (143547 bp).

Discussion

We began this study with the assumption that there would be a higher chance of finding a phage and host pair if both came from the same environment. No phages were isolated using the same source for bacterial host and phage isolation, and only one was found using a bank of laboratory strains (Fig. 1). This lack of success is not unprecedented. In the isolation of DW2 and CS1 only five out of twenty different farmyard slurry samples gave plaques against *S. aureus* DPC 5246 (13). Restriction analysis indicated that only two distinct phages were present, DW2 and CS1. In a study of the isolation of phages from silage on dairy farms 114 *Listeria* phages were isolated (16). Based on host range and genome size re-isolation of the same phages was indicated. Vongkamjan, *et al*. also noted that certain serotypes of *Listeria* acted as

hosts more often than others suggesting that bacterial choice for screening is important. Only nine phages were isolated when screening 86 waste water samples against 88 *S. aureus* strains (17). Six were subsequently excluded due to their lysogenic lifecycle or low lytic activity. During the isolation of phages for phage therapy against clinically relevant pathogens only one phage was isolated against one *S. aureus* strain after 117 enrichment attempts (18). No phages were isolated against the other nine *S. aureus* strains used. Multiple phages were isolated from the same source but these were judged to be the same based on the plaque appearance. Mattila, *et al*. also found that the probability of isolating a phage varied based on the pathogen of interest (18). Phage isolation was successful almost 90% of the time against *Salmonella* strains but only 30% of the time against *Enterococcus faecium*/*faecalis* strains. This can occur if the host that the phage is isolated against is not be the most effective or most recent natural host (19). To date only temperate phages, but no virulent phages, have been isolated against *Clostridium difficile,* the causative agent of *C. difficile* infection (20). Similarly, no virulent *Bifidobacterium* phages have been isolated although a number of temperate phages have been identified (21). Therefore, the choice of host is important for phage isolation and the isolation of the same phages multiple times is common.

The inability to isolate phages from the same environment that their host was found in may be due to how we isolate phages. A number of things must be taken in to account when screening for phages. Phage screening can be carried out using direct plating or enrichment before plating (22). Samples can be concentrated by filtration or precipitation before screening. Direct plating and enrichment have both been effective in isolating phages. Direct plating requires higher concentrations of phage particles in the samples being screened than samples being used for enrichment. Enrichment can

negatively bias isolation towards faster growing phages or phages that propagate better in liquid culture and miss slower growing phages or phages that perform better in environmental situations. Conversely, this can act as positive selection for the fastest and fittest phages. Enrichment can be effective in increasing the number of phages isolated from a source. By using enrichment Jurczak-Kurek, *et al*. isolated 83 potential phages from two sewage samples from a wastewater treatment plant (23). Using morphological and physiological analysis 13 phages were determined to be duplications and discounted. Concentration of samples can be carried out before direct plating to give a more accurate representation of an environment than enrichment. Faster growing or less pathogenic "surrogate" strains can be used for screening to make the process faster and easier (24). These surrogates can also be used solely for propagation rather than isolation. Obviously, no phages can be isolated against unculturable bacteria. Most phages have been isolated using a single host but the use of multiple hosts combined can aid in the isolation of broader host range phages (22).

There are a number of negatives associated with the most common methods for investigating phages such as plaque and spot assays. They can be labour intensive, the bacterial hosts may not form confluent lawns, spot assays can show false positives due to phage binding but no infection (lysis from without), and plaque assays can show false negatives due to incorrect media or low infection productivity. Plaque formation can be media dependent. For example, jumbo phages, with genomes greater than 200000 bp and large phage virions form small plaques in agarose concentrations commonly used in screening (25). These small plaques may not be visible. The large virion size does not allow for diffusion through higher agarose concentrations so lower concentrations must be used for larger, visible plaques to form. Jumbo phage have also been found to not lyse liquid bacterial cultures like smaller phages.

Not finding a phage in the same source as its host may have also been indicative of phage and bacterial community dynamics. Phage and bacterial numbers in a community are dependent on one another with phages helping to maintain bacterial diversity (26). Areas of the human gut, for example the lumen and mucosa, can represent different environments and different dynamics models can be present within them. Phages may not have been isolated as their numbers were low. Phages present in high numbers would have been available for isolation but their bacterial hosts perhaps were only present in low numbers so would not be isolated.

When the same faecal sample was screened for phages against a bank of 77 known strains only one phage was isolated and it was determined to be a previously sequenced and described phage, DW2 (13) (Fig. 2). In our study, DW2 and CS1 had very similar host ranges (Table 3). This was in contrast to O'Flaherty, *et al*. who found DW2 and CS1 to have some similarities in host range but also to infect some different hosts and have differing EOPs. SARL was not detected in the virome of 12 monthly samples of the same subject. However, the longitudinal study took place over a year after SARL was isolated from the subject. It is possible that it was present but at very low abundance and the metagenome coverage was too low to detect it. Another possibility is that it was not replicating in the gut and was simply passing through and was of food or environmental origin (27). Or it too may have been due to community dynamics as mentioned in relation to the isolation of phages and their bacterial hosts from the same source.

A number of previously identified phages were re-isolated in this study (Fig. 3). Phage JL1 is a 43,457 bp *Siphoviridae* phage first isolated against *Escherichia coli* O157:H7 (Accession Number JX865427) (28). JL1 was isolated from sewage. Phage vB_EcoM_JS09 is a 169,148 bp *Myoviridae* phage first isolated against

enterotoxigenic *E. coli* (Accession Number KF582788) (29). vB_EcoM_JS09 was isolated from sewage samples from a pig farm. Phage vB_Kpn_IME260 is a 123,490 bp phage first isolated against *Klebsiella pneumoniae* (Accession Number KX845404) (30). vB_Kpn_IME260 was isolated from sewage water. Phage OSYSP is a 110,901 bp phage first isolated against *E. coli* O157:H7 from waste water (Accession Number MF402939) (31). Phage Pet-CM3-4 is a 171, 975 bp *Myoviridae* phage first isolated against *Cronobacter malonaticus* (Accession Number LT614807). Pet-CM3-4 was isolated from waste water.

APCEc01 was first isolated from the faecal sample, used in a metagenomic study, of an elderly patient in long-term care with no known health disorders (Accession Number KR422352) (15). Two other phages, APCEc02 and APCEc03, were isolated from other individuals in the same metagenomic study. APCEc01 was isolated from a faecal sample with a higher *E. coli/Shigella* content than the samples from which APCEc02 and APCEc03 were isolated, 2% compared to 0.07% and 0.24%, respectively. APCEc01 is a 168,771 bp *Myoviridae* phage first isolated against *E. coli* strain APC 106. No genes suggesting a temperate lifestyle were found. APCEc01 applied for 24 hours at a concentration higher than 10^7 PFU/well reduced biofilm activity by at least 4-fold. APCEc01 was considered to have a narrow host range in comparison to the other two phages isolated at the same time, targeting only three of the 15 *E. coli* strains tested compared to APCEc02 which targeted five strains and APCEc03 which targeted nine strains. APCEc01 infected *E. coli* strains APC 104, APC 105, and APC 106 and a *Shigella sonnei* strain isolated from a patient with shigellosis. This ability to infect more than one species may help to explain why APCEc01 was such a dominant phage and was found so many times in our experiment. *E. coli* is commonly found in farm environments which could also have explained the

number of APCEc01 strains isolated (15). The samples of maize silage, grass silage, cow faeces, and sheep faeces were from the same farm so APCEc01 could have been transferred between these places and contributed towards its presence in multiple areas. APC 104, APC 105, and APC 106 were included for phage screening in this study. The GC content of APCEc01 at 37.7% is lower than that of *E. coli*, which is around 50%, suggesting that elements of the genome may have been acquired from phages infecting hosts with a lower GC content or could have evolved from an ancestor that infected a host with a lower GC content (15). APCEc01 and vB_EcoM_JS09 have been placed in the same clade, *Rb69virus* (32). SPAdes is commonly used in phageomic studies and forms consensus sequences for very similar sequences and can mean that the variation between individual strains is lost. For our analysis all of the individual strain sequences of APCEc01 isolated in this study were used.

Different phages show different rates of mutation or evolution. In a metagenomics study of the human gut virome over a 2.5 year period temperate phages showed lower mutation rates, consistent with replication by accurate bacterial DNA polymerases in the integrated prophage state, than lytic phages (33). The single stranded DNA phage belonging to the *Microviridae* showed high substitution rates and up to 4% variation over a 2.5 year period. Substitutions occurred at a steady rate as well as in response to a change in selective pressure. *Podoviridae*, *Myoviridae*, and *Siphoviridae* had medium substitution rates. This rapid evolution was suggested as a contributor to the inter-individual nature of human gut viromes. The International Committee on Taxonomy of Viruses defines different *Microviridae* species as having as little as 3.1% divergence. Two viruses belonging to the same species differ from

each other by less than 5% at the nucleotide level according to the International Committee on the Taxonomy of Viruses (4).

Synonymous SNPs exhibit extremely low mutation rates and are expected to be evolutionarily neutral or near so, therefore providing useful markers for studying population genetics and the epidemiology of bacterial pathogens (34). For genes in which only synonymous SNPs were found, such as topoisomerase and RNA ligase, it could be possible that non-synonymous changes would negatively affect their expression or operation and so only synonymous SNPs have survived (Fig. 7). Not all phage genes undergo mutation at the same rate (19). This was visible when SNP counts were normalised based on gene length (Fig. 7). The head genes do not often undergo horizontal gene transfer (17). The proteins produced by the head genes must interact in a complex way during the formation of new phages. Newly evolved genes may produce proteins that cannot interact adequately. This does not mean that these genes undergo less mutation it simply means that in these areas mutations that are unsuccessful or non-functioning are selected against. A phage baseplate wedge subunit was found to have low levels of only synonymous changes which could have occurred for similar reasons to phage head genes undergoing low levels of horizontal gene transfer (Fig. 7). A phage rIIA lysis inhibitor gene was found to have only non-synonymous SNPs and in one case this introduced a stop codon. Similarly, the introduction of stop codons in genes could be indicative that these genes can be altered without affecting the function of the proteins they encode. There is also the possibility that these introduced stop codons, or those in the original genome, will not function due to processes such as stop codon readthrough in which stop codons function as sense codons (35).

SNPs can have important effects on physiology. Non-core phage genes, which are not shared between all members of a group, can aid in the exploitation of a niche. For example, temperate phages can use the diversity-generating retroelements (DGR) system to diversify tail fibres (36). This can be used to overcome bacterial resistance mechanisms, such as the mutation of bacterial surface receptors, by introducing up to 10^{14} tail fibre variations. A change in the class of the amino acid is more likely to affect function than a change to the same class (37). A phenotypic change may require the interaction of a number of genetic changes and environmental conditions. Phages experimentally coevolved with *P. fluorescens* increased host range over time and those with the broadest host range were found to have the most non-synonymous changes (38). These were mostly located in the phage tail fibre gene. It was suggested that these non-synonymous mutations had an epistatic effect as the changes could not be linked directly to infection of a specific host. The effect of competition and selective pressure was evident in the study by Scanlan, *et al*. where phages that were evolved in the presence of non-evolving hosts did not show mutations in select genes while those in the presence of evolving phages did (38). Evolution of phage infectivity is a combination of phage genomics and host sensitivity. Conversely, phages can also evolve a reduced host range (39). Phages have been found to increase their host range with as few as one to three mutations in target genes (40). The host range or growth of the individual APCEc01 strains was not assessed in this study but it would be useful to carry this out in the future and to see if any substitutions changed any aspects of the phage growth.

Although many phage genes have been annotated the function of many is still unknown (41). SNPs may be present in a gene we class as non-coding or of unknown function but that is simply because we have not established its role yet. This was

evident in the data presented here where a number of SNPs occurred in hypothetical proteins or genes with ambiguous names such as phage protein for which no function has been determined (Fig. 7). Similarly, a number of hypothetical proteins were found to have no SNPs (Fig. 6). We cannot predict why this occurred if no function can be assigned. Therefore, the effect of SNPs in these areas cannot be predicted or even suggested highlighting the need for greater basic phage research.

It could be possible that SNPs occurred in phage regions targeted by the CRISPR/Cas system of its host. SNPs in the proto-spacer or the proto-spacer adjacent motif (PAM) of phages can allow phages to overcome CRISPR/Cas associated immunity (42). The PAM is a conserved sequence immediately next to the protospacer that is used by the bacteria to discriminate between its own nucleic acid and that of invading phages. This could explain why SNPs were found in blocks and not equally dispersed throughout the genome. It could also explain why in a number of cases the exact same SNP was found in a number of strains as phages with PAM mutations were found to be dominant over wild type phages (42). Coevolution can occur by the reciprocal acquisition of new CRISPR spacers by the bacterial host and mutation in these regions by the phage (43). It would be interesting to look if the SNPs in the APCEc01 genome are in regions targeted by the CRISPR/Cas system.

One aspect of this study concentrated on the importance of bacterial strain selection and environmental source for screening for phages. It drew attention to the difficulty in isolating new and unreported phages. A number of previously known phages were re-isolated together with a single novel phage. The second part of this work looked at the occurrence of SNPs in strains of a single phage isolated from various environmental sources. Up to 800 SNPs occurred between some genomes including instances where stop codons were introduced and amino acids were

changed. It was also significant to see where SNPs did and did not occur which can highlight the importance of these genes or their requirement to be conserved. The large number of hypothetical proteins and genes known only as phage proteins also showed how much is left to learn about the genomes of phages. These changes could be significant considering the effect that small genetic changes can have on phage and bacterial metabolism.

References

1. **Rozin P, Spranca M, Krieger Z, Neuhaus R, Surillo D, Swerdlin A, Wood K.** 2004. Preference for natural: instrumental and ideational/moral motivations, and the contrast between foods and medicines. Appetite **43**:147-154.

2. **Coffey B, Mills S, Coffey A, McAuliffe O, Ross RP.** 2010. Phage and their lysins as biocontrol agents for food safety applications. Annu Rev Food Sci Technol **1:**449-68.

3. **Gordillo Altamirano FL, Barr JJ.** 2019. Phage Therapy in the Postantibiotic Era. Clinical Microbiology Reviews **32**:e00066-18.

4. **Adriaenssens E, Brister JR.** 2017. How to Name and Classify Your Phage: An Informal Guide. Viruses **9**:70.

5. **Abedon ST.** 2009. Chapter 1 Phage Evolution and Ecology, p 1-45, Advances in Applied Microbiology, vol **67**. Academic Press.

6. **Brüssow H, Desiere F.** 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol Microbiol **39:**213-22.

7. **Buckling A, Rainey PB.** 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society B: Biological Sciences **269:**931-936.

8. **Vollmers J, Wiegand S, Kaster A-K.** 2017. Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! PLOS ONE **12**:e0169662.

9. **Koskella B, Meaden S.** 2013. Understanding Bacteriophage Specificity in Natural Microbial Communities. Viruses **5:**806-823.

10. **Villarroel J, Kleinheinz KA, Jurtz VI, Zschach H, Lund O, Nielsen M, Larsen MV.** 2016. HostPhinder: A Phage Host Prediction Tool. Viruses **8**.

11. **Duffy S, Turner PE, Burch CL.** 2006. Pleiotropic costs of niche expansion in the RNA bacteriophage phi 6. Genetics **172**:751-757.

12. **Keary R, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A.** 2014. Genome analysis of the staphylococcal temperate phage DW2 and functional studies on the endolysin and tail hydrolase. Bacteriophage **4**:e28451.

13. **O'Flaherty S, Ross RP, Flynn J, Meaney WJ, Fitzgerald GF, Coffey A.** 2005. Isolation and characterization of two anti-staphylococcal bacteriophages specific for pathogenic *Staphylococcus aureus* associated with bovine infections. Lett Appl Microbiol **41:**482-6.

14. **Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, McDonnell SA, Khokhlova EV, Draper LA, Forde A, Guerin E, Velayudhan V, Ross RP, Hill C.** 2019. The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. Cell Host & Microbe **26**:527-541.e5.

15. **Dalmasso M, Strain R, Neve H, Franz CMAP, Cousin FJ, Ross RP, Hill C.** 2016. Three New *Escherichia coli* Phages from the Human Gut Show Promising Potential for Phage Therapy. PLOS ONE **11**:e0156773.

16. **Vongkamjan K, Switt AM, den Bakker HC, Fortes ED, Wiedmann M.** 2012. Silage collected from dairy farms harbors an abundance of listeriaphages with considerable host range and genome size diversity. Applied and environmental microbiology **78**:8666-8675.

17. **Łubowska N, Grygorcewicz B, Kosznik-Kwaśnicka K, Zauszkiewicz-Pawlak A, Węgrzyn A, Dołęgowska B, Piechowicz L.** 2019. Characterization of the Three New Kayviruses and Their Lytic Activity Against Multidrug-Resistant Staphylococcus aureus. Microorganisms **7**:471.

18. **Mattila S, Ruotsalainen P, Jalasvuori M.** 2015. On-Demand Isolation of Bacteriophages Against Drug-Resistant Bacteria for Personalized Phage Therapy. Frontiers in microbiology **6:**1271-1271.

19. **Hatfull GF, Hendrix RW.** 2011. Bacteriophages and their Genomes. Current opinion in virology **1**:298-303.

20. **Hargreaves KR, Clokie MRJ.** 2014. *Clostridium difficile* phages: still difficult? Frontiers in Microbiology **5**.

21. **Mahony J, Lugli GA, van Sinderen D, Ventura M.** 2018. Impact of gut-associated bifidobacteria and their phages on health: two sides of the same coin? Applied Microbiology and Biotechnology **102**:2091-2099.

22. **Hyman P.** 2019. Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. Pharmaceuticals (Basel, Switzerland) **12**:35.

23. **Jurczak-Kurek A, Gąsior T, Nejman-Faleńczyk B, Bloch S, Dydecka A, Topka G, Necel A, Jakubowska-Deredas M, Narajczyk M, Richert M, Mieszkowska A, Wróbel B, Węgrzyn G, Węgrzyn A.** 2016. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. Scientific Reports **6:**34338.

24. **Gill JJ, Hyman P.** 2010. Phage choice, isolation, and preparation for phage therapy. Curr Pharm Biotechnol **11:**2-14.

25. **Serwer P, Hayes SJ, Thomas JA, Hardies SC.** 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. Virology journal **4**:21-21.

26. **Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP.** 2013. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes **4**:4-16.

27. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.** 2011. The human gut virome: Inter-individual variation and dynamic response to diet. Genome Research **21**:1616-1625.

28. **Pan F, Wu H, Liu J, Ai Y, Meng X, Meng R, Meng Q.** 2013. Complete genome sequence of *Escherichia coli* O157:H7 lytic phage JL1. Archives of Virology **158:**2429-2432.

29. **Zhou Y, Bao H, Zhang H, Wang R.** 2015. Isolation and Characterization of Lytic Phage vB_EcoM_JS09 against Clinically Isolated Antibiotic-Resistant Avian Pathogenic *Escherichia coli* and Enterotoxigenic *Escherichia coli*. Intervirology **58:**218-31.

30. **Xing S, Pan X, Sun Q, Pei G, An X, Mi Z, Huang Y, Zhao B, Tong Y.** 2017. Complete Genome Sequence of a Novel Multidrug-Resistant *Klebsiella pneumoniae* Phage, vB_Kpn_IME260. Genome announcements **5**:e00055-17.

31. **Yesil M, Huang E, Yang X, Yousef AE.** 2017. Complete Genome Sequence of Escherichia Phage OSYSP. Genome announcements **5**:e00880-17.

32. **Michniewski S, Redgwell T, Grigonyte A, Rihtman B, Aguilo-Ferretjans M, Christie-Oleza J, Jameson E, Scanlan DJ, Millard AD.** 2019. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. Environ Microbiol **21**:2112-2128.

33. **Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD.** 2013. Rapid evolution of the human gut virome. Proceedings of the National Academy of Sciences **110**:12450-12455.

34. **Zhang W, Qi W, Albert TJ, Motiwala AS, Alland D, Hyytia-Trees EK, Ribot EM, Fields PI, Whittam TS, Swaminathan B.** 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. Genome Research **16**:757-767.

35. **Wenthzel A-MK, Stancek M, Isaksson LA.** 1998. Growth phase dependent stop codon readthrough and shift of translation reading frame in *Escherichia coli*. FEBS Letters **421**:237-242.

36. **Ofir G, Sorek R.** 2018. Contemporary Phage Biology: From Classic Models to New Insights. Cell **172**:1260-1270.

37. **Guard J, Morales CA, Fedorka-Cray P, Gast RK.** 2011. Single nucleotide polymorphisms that differentiate two subpopulations of *Salmonella* enteritidis within phage type. BMC Research Notes **4:**369.

38. **Scanlan PD, Hall AR, Lopez-Pascua LD, Buckling A.** 2011. Genetic basis of infectivity evolution in a bacteriophage. Mol Ecol **20**:981-9.

39. **Duffy S, Burch CL, Turner PE.** 2007. Evolution of host specificity drives reproductive isolation among RNA viruses. Evolution **61**:2614-22.

40. **Hall Alex R, Scanlan Pauline D, Buckling A.** 2011. Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens. The American Naturalist **177**:44-53.

41. **Sutton TDS, Clooney AG, Ryan FJ, Ross RP, Hill C.** 2019. Choice of assembly software has a critical impact on virome characterisation. Microbiome **7**:12.

42. **Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF.** 2013. Phage mutations in response to CRISPR diversification in a bacterial population. Environ Microbiol **15:**463-70.

43. **Common J, Morley D, Westra ER, van Houte S.** 2019. CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage. Philos Trans R Soc Lond B Biol Sci **374**:20180098.

44. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA.** 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res **27**:824-834.

45. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215**:403-10.

46. **González-Tortuero E, Sutton TDS, Velayudhan V, Shkoporov AN, Draper LA, Stockdale SR, Ross RP, Hill C.** 2018. VIGA: a sensitive, precise and automatic *de novo* VIral Genome Annotator. bioRxiv doi:10.1101/277509:277509.

47. **Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G.** 2010. Interactive microbial genome visualization with GView. Bioinformatics **26:**3125-6.

Tables and figures

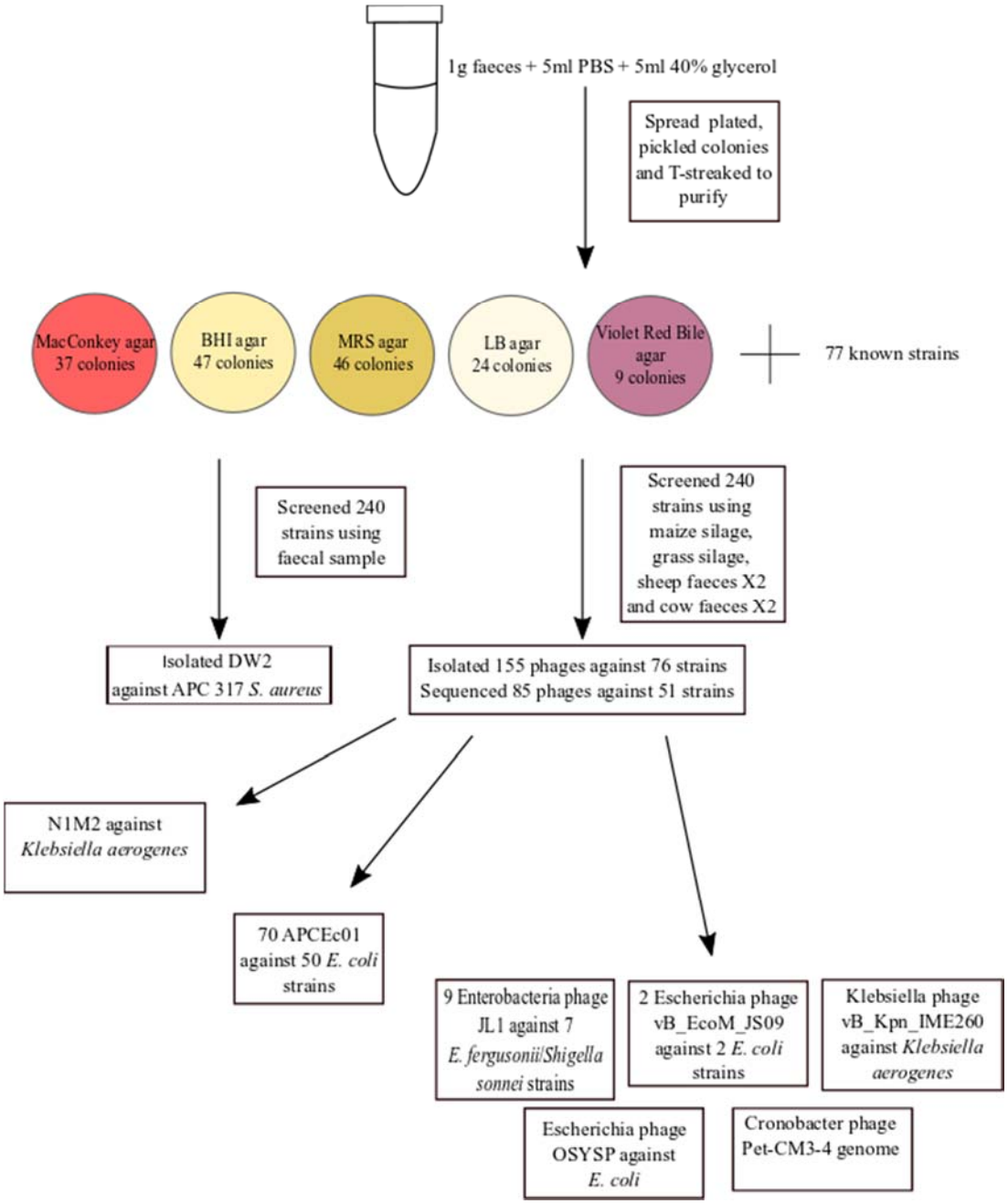


Figure 1. Workflow of bacterial strain isolation from faeces and screening of faecal and environmental (maize silage, grass silage, sheep faeces, and cow faeces) samples for phages.

Table 1. Phages isolated and sequenced. Bacteria in brackets are laboratory strains and were not isolated in this study. Phages were named using the convention "A/N-bacterial host strain number-M/G/S1/S2/C1". A=aerobic isolation, N=anaerobic isolation. M=maize silage, G=grass silage, S1=sheep faeces sample 1, S2=sheep faeces sample 2, C1=cow faeces sample 1. For some samples an extra number was added to the end if multiple phages were isolated from the same source against the same bacterial host.

Phage	Bacterial host 16S	Media host isolated from	Phage annotation
A1G1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A1G2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A1M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A1M2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A1M3	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A4G1	*Shigella sonnei/Shigella flexneri/E.fergusonii*	MacConkey	Not sequenced
A4G2	*Shigella sonnei/Shigella flexneri/E.fergusonii*	MacConkey	APCEc01
A4M	*Shigella sonnei/Shigella flexneri/E.fergusonii*	MacConkey	APCEc01
A5G	*Shigella sonnei/Shigella flexneri/E.fergusonii*	MacConkey	Not sequenced
A5M1	*Shigella sonnei/Shigella flexneri/E.fergusonii*	MacConkey	APCEc01
A5M2	*Shigella sonnei/Shigella flexneri/E.fergusonii*	MacConkey	Not sequenced
A7G	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
A7M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
A8G	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	Not sequenced
A8M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
A10G	Not sequenced	BHI	Cronobacter phage Pet-CM3-4
A21S2	*E.marmotae/E.fergusonii/Shigella boydii*	MacConkey	Not sequenced
A22M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A22M2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A22S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A24M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A24S21	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A24S22	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A25M	Not sequenced	MacConkey	APCEc01
A25S21	Not sequenced	MacConkey	APCEc01
A25S22	Not sequenced	MacConkey	Not sequenced
A25S23	Not sequenced	MacConkey	APCEc01
A26M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A26M2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A27M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	Not sequenced
A27S21	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A27S22	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A30M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
A35M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MRS	APCEc01
A35S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MRS	Not sequenced
A37M	*E.fergusonii/Shigella boydii/E.fergusonii*	MRS	Not sequenced
A41M	*E.fergusonii/Shigella sonnei/E.fergusonii*	MRS	Not sequenced
A41S2	*E.fergusonii/Shigella sonnei/E.fergusonii*	MRS	Enterobacteria phage JL1
A43M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MRS	APCEc01
A43S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MRS	Not sequenced
A44M	*E.fergusonii/Shigella boydii/Shigella dysenteriae*	MRS	Not sequenced

A44S2	*E.fergusonii/Shigella boydii/Shigella dysenteriae*	MRS	Not sequenced
A45M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MRS	APCEc01
A46S2	*E.fergusonii/Shigella boydii/Shigella dysenteriae*	MRS	Not sequenced
A50M	*Shigella sonnei/Shigella flexneri/Shigella dysenteriae*	LB	APCEc01
A50S2	*Shigella sonnei/Shigella flexneri/Shigella dysenteriae*	LB	APCEc01
A51S2	*E.fergusonii/Shigella sonnei/E.fergusonii*	LB	Enterobacteria phage JL1
A53M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	LB	APCEc01
A53S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	LB	Not sequenced
A59M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	APCEc01
A59M2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	APCEc01
A59S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	APCEc01
A61M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	APCEc01
A61M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	Not sequenced
A61S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	APCEc01
A66M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	Not sequenced
A66S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	VRBA	Not sequenced
A67C11	*E.fergusonii/Shigella sonnei/E.fergusonii*	VRBA	Enterobacteria phage JL1
A67C12	*E.fergusonii/Shigella sonnei/E.fergusonii*	VRBA	Enterobacteria phage JL1
A67S2	*E.fergusonii/Shigella sonnei/E.fergusonii*	VRBA	Enterobacteria phage JL1
A68M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
A68S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
A70M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	Not sequenced
A70S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
A71M1	*E.fergusonii/Shigella boydii/E.fergusonii*	BHI	Not sequenced
A71M2	*E.fergusonii/Shigella boydii/E.fergusonii*	BHI	APCEc01
A73M	Not sequenced	MRS	Not sequenced
A73S2	Not sequenced	MRS	Not sequenced
A74S2	*E.fergusonii/Shigella boydii/Shigella sonnei*	MRS	APCEc01
A77M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MRS	Not sequenced
A77S2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MRS	APCEc01
A78M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	APCEc01
A78S2	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	Not sequenced
A81G	Not sequenced	BHI	Not sequenced
A81M	Not sequenced	BHI	Not sequenced
A81S11	Not sequenced	BHI	Not sequenced
A81S12	Not sequenced	BHI	Not sequenced
A85M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	APCEc01
APC104G	(*E. coli* APC104)		APCEc01
APC105G	(*E. coli* APC105)		APCEc01
APC105M1	(*E. coli* APC105)		APCEc01
APC105M2	(*E. coli* APC105)		Not sequenced
APC106G1	(*E. coli* APC106)		Not sequenced
APC106G2	(*E. coli* APC106)		Not sequenced
APC106M	(*E. coli* APC106)		APCEc01
MG1655G	(*E. coli* MG1655)		APCEc01
MG1655M1	(*E. coli* MG1655)		Not sequenced
MG1655M2	(*E. coli* MG1655)		Escherichia phage OSYSP
SARL	(*S. aureus* APC317)		DW2
N1M1	*Klebsiella aerogenes*	MacConkey	Klebsiella phage vB_Kpn_IME260
N1M2	*Klebsiella aerogenes*	MacConkey	Novel N1M2
N9G	*E.fergusonii/Shigella flexneri/Shigella sonnei*	BHI	APCEc01
N9M1	*E.fergusonii/Shigella flexneri/Shigella sonnei*	BHI	APCEc01
N9M2	*E.fergusonii/Shigella flexneri/Shigella sonnei*	BHI	Not sequenced
N17G	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
N17M	*Shigella sonnei/E.fergusonii/Shigella flexneri*	MacConkey	APCEc01
N18M	Not sequenced	MacConkey	Not sequenced
N19G1	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MacConkey	Not sequenced
N19G2	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MacConkey	Not sequenced
N22G1	*E.fergusonii/Shigella flexneri/Shigella sonnei*	MacConkey	Not sequenced
N22G2	*E.fergusonii/Shigella flexneri/Shigella sonnei*	MacConkey	Not sequenced

N22M	*E.fergusonii/Shigella flexneri/Shigella sonnei*	MacConkey	APCEc01
N26M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MacConkey	APCEc01
N27M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MacConkey	APCEc01
N28G	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MacConkey	APCEc01
N28M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	MacConkey	Not sequenced
N29G	*E.fergusonii/Shigella boydii/Shigella sonnei*	MacConkey	Enterobacteria phage JL1
N30M	*E.fergusonii/Shigella boydii/Shigella sonnei*	MacConkey	Not sequenced
N31G	*E.fergusonii/Shigella flexneri/E.coli*	MacConkey	Enterobacteria phage JL1
N32M	*E.fergusonii/Shigella boydii/Shigella flexneri*	MacConkey	APCEc01
N33M	*E.fergusonii/Shigella boydii/Shigella flexneri*	MacConkey	APCEc01
N36M	*E.fergusonii/Shigella boydii/Shigella flexneri*	BHI	APCEc01
N37M1	*Shigella flexneri/E.fergusonii/E.fergusonii*	BHI	APCEc01
N37M2	*Shigella flexneri/E.fergusonii/E.fergusonii*	BHI	Not sequenced
N38G	Not sequenced	BHI	Enterobacteria phage JL1
N39G	*E.fergusonii/Shigella flexneri/Shigella sonnei*	BHI	Not sequenced
N41M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	APCEc01
N43M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	Not sequenced
N48G	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	APCEc01
N48M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	BHI	Not sequenced
N50G	*E.fergusonii/E.marmotae/Shigella boydii*	MRS	APCEc01
N50M	*E.fergusonii/E.marmotae/Shigella boydii*	MRS	Not sequenced
N51M	*Shigella flexneri/E.fergusonii/E.fergusonii*	MRS	APCEc01
N54G	Not sequenced	MRS	Not sequenced
N54M1	Not sequenced	MRS	Not sequenced
N54M2	Not sequenced	MRS	Not sequenced
N55G	*E.fergusonii/Shigella flexneri/Shigella sonnei*	MRS	APCEc01
N58G	*Shigella flexneri/E.fergusonii/E.fergusonii*	MRS	Enterobacteria phage JL1
N58M1	*Shigella flexneri/E.fergusonii/E.fergusonii*	MRS	Not sequenced
N58M2	*Shigella flexneri/E.fergusonii/E.fergusonii*	MRS	Not sequenced
N59M1	Not sequenced	MRS	Not sequenced
N59M2	Not sequenced	MRS	Not sequenced
N62G	*Shigella flexneri/E.fergusonii/E.fergusonii*	LB	Escherichia phage vB_EcoM_JS09
N62M1	*Shigella flexneri/E.fergusonii/E.fergusonii*	LB	Not sequenced
N62M2	*Shigella flexneri/E.fergusonii/E.fergusonii*	LB	Not sequenced
N64G	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	Not sequenced
N64M	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	APCEc01
N65G1	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	Not sequenced
N65G2	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	APCEc01
N65G3	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	Not sequenced
N65M1	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	APCEc01
N65M2	*Shigella flexneri/Shigella sonnei/E.fergusonii*	LB	APCEc01
N66M1	*E.fergusonii/Shigella flexneri/Shigella sonnei*	LB	APCEc01
N66M2	*E.fergusonii/Shigella flexneri/Shigella sonnei*	LB	Not sequenced
N67G	*Shigella sonnei/E.fergusonii/Shigella flexneri*	LB	Not sequenced
N67M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	LB	APCEc01
N67M2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	LB	Not sequenced
N67M3	*Shigella sonnei/E.fergusonii/Shigella flexneri*	LB	APCEc01
N68G	*Shigella dysenteriae/Shigella sonnei/Shigella flexneri*	LB	Escherichia phage vB_EcoM_JS09
N68M	*Shigella dysenteriae/Shigella sonnei/Shigella flexneri*	LB	Not sequenced
N69G	*Shigella sonnei/Shigella flexneri/E.fergusonii*	LB	APCEc01
N69M1	*Shigella sonnei/Shigella flexneri/E.fergusonii*	LB	APCEc01
N69M2	*Shigella sonnei/Shigella flexneri/E.fergusonii*	LB	Not sequenced
N73G	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01
N73M1	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	Not sequenced
N73M2	*Shigella sonnei/E.fergusonii/Shigella flexneri*	BHI	APCEc01

Table 2. A bank of known strains was selected to be used for phage screening.

Strain designation	Other designations	Strain
APC 53	DPC 6085, NCIMB700577, NCDO0577	*Bacillus cereus*
APC 54	DPC 6086, NCIMB700578, NCDO0578	*Bacillus cereus*
APC 55	DPC 6087, NCIMB700579, NCDO0579	*Bacillus cereus*
APC 56	DPC 6088, NCIMB700827, NCDO0827	*Bacillus cereus*
APC 57	DPC 6089, NCIMB8079, ATCC7004	*Bacillus cereus*
APC 58	DPC 6334, Bel 17Bc	*Bacillus cereus*
APC 59	DPC 6336, Bel Bc33	*Bacillus cereus*
DPC 8079		*Bacillus cereus*
APC 1762	UCC 5002	*Bacillus subtilis*
APC 1640		*Enterococcus casseliflavus*
APC 1039	EC618	*Enterococcus faecalis*
APC 1749	DPC 5152	*Enterococcus faecalis*
APC 1025	EC251	*Enterococcus faecium*
APC 1026	EC289	*Enterococcus faecium*
APC 1029	EC300	*Enterococcus faecium*
APC 1030	EC357	*Enterococcus faecium*
APC 1031	EC520	*Enterococcus faecium*
APC 1032	EC533	*Enterococcus faecium*
APC 1035	EC548	*Enterococcus faecium*
APC 1036	EC562	*Enterococcus faecium*
APC 1038	EC587	*Enterococcus faecium*
APC 1043	*EC725*	*Enterococcus faecium*
APC 1044	*EC748*	*Enterococcus faecium*
APC 1641		*Enterococcus faecium*
Lactobacillus salivarius 6482		*Lactobacillus salivarius*
APC 317	DPC 5245	*S. aureus*
APC 993	ST290	*S. aureus*
APC 994	ST291	*S. aureus*
APC 995	ST295	*S. aureus*
APC 996	ST299	*S. aureus*
APC 998	ST355	*S. aureus*
APC 1003	ST528	*S. aureus*
APC 1004	ST530	*S. aureus*
APC 1007	ST535	*S. aureus*
APC 1009	ST544	*S. aureus*
APC 1010	ST550	*S. aureus*
APC 1012	35197	*S. aureus*
APC 1019	25949	*S. aureus*
APC 1022	Newman	*S. aureus*
APC 1024	RF122	*S. aureus*
APC 1055	LMG14694, ATCC 13813, CCRC 10787, CCUG 4208, CIP 103227, DSM 2134, JCM 5671, NCFB 1348, NCTC 8181	*Streptococcus agalactiae*
APC 1759		*Streptococcus agalactiae*
APC 1755		*Streptococcus dysgalactiae*
APC 119	DPC 6143, 4001	*Streptococcus mutans*
APC 120	DC 6144, 4021	*Streptococcus mutans*
APC 121	DPC 6145, 4070	*Streptococcus mutans*
APC 122	DPC 6150, 4009	*Streptococcus mutans*
APC 123	DPC 6151, 4030	*Streptococcus mutans*
APC 124	DPC 6152, 4037	*Streptococcus mutans*
APC 125	DPC 6153, 4039	*Streptococcus mutans*
APC 126	DPC 6154, 4040	*Streptococcus mutans*

APC 127	DPC 6155, 4055	*Streptococcus mutans*
APC 128	DPC 6156, 4058	*Streptococcus mutans*
APC 129	DPC 6157, 3007	*Streptococcus mutans*
APC 130	DPC 6158, 3013	*Streptococcus mutans*
APC 131	DPC 6159, 3017	*Streptococcus mutans*
APC 132	DPC 6160, 1038	*Streptococcus mutans*
APC 133	DPC 6161, 1054	*Streptococcus mutans*
APC 134	DPC 6162, NCTC10449, ATCC 25175, SIMS, DSM 20523, ATCC 25175, IFO 13955, NCDO 2062	*Streptococcus mutans*
APC 135	DPC 6543	*Streptococcus mutans*
APC 1756	DSM 2071	*Streptococcus pyogenes*
APC 1757	DSM11728	*Streptococcus pyogenes*
APC 1758	NCDO2381	*Streptococcus pyogenes*
APC 104	DPC 6009	*Escherichia coli*
APC 105	DPC 6050	*Escherichia coli*
APC 106	DPC 6051	*Escherichia coli*
APC 109	DPC 6054, P1432	*Escherichia coli*
APC 110	DPC 6055, AR12900	*Escherichia coli*
APC 115	DPC 6472	*Escherichia coli*
APC 1220	HM605	*Escherichia coli*
Escherichia coli 042		*Escherichia coli*
Escherichia coli Nissle		*Escherichia coli*
Esherichia coli UTI89		*Escherichia coli*
APC 1977	MG1655	*Escherichia coli*
APC 176	DPC 6452	*Salmonella typhimurium*
APC 181	DPC 6547	*Salmonella typhimurium*
APC 185	DPC 6436	*Salmonella typhimurium*

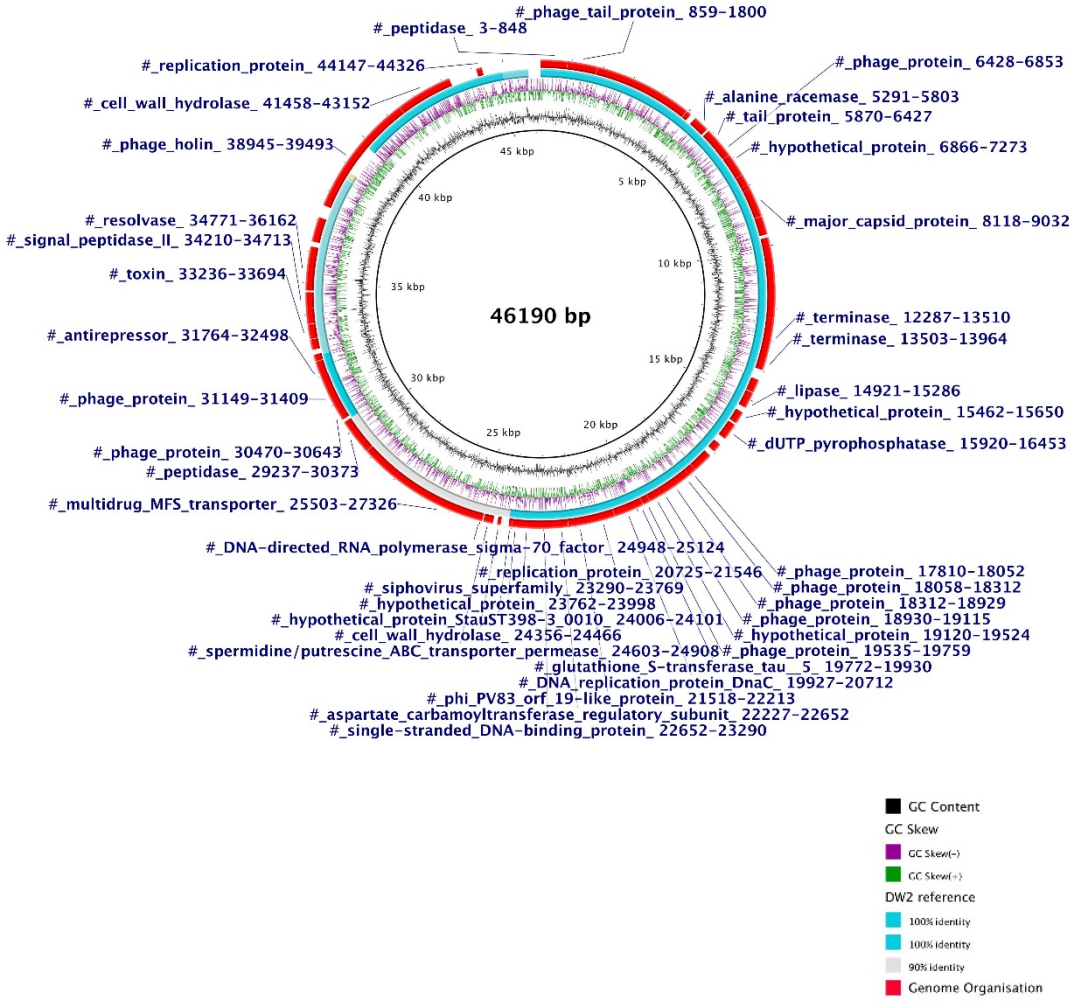
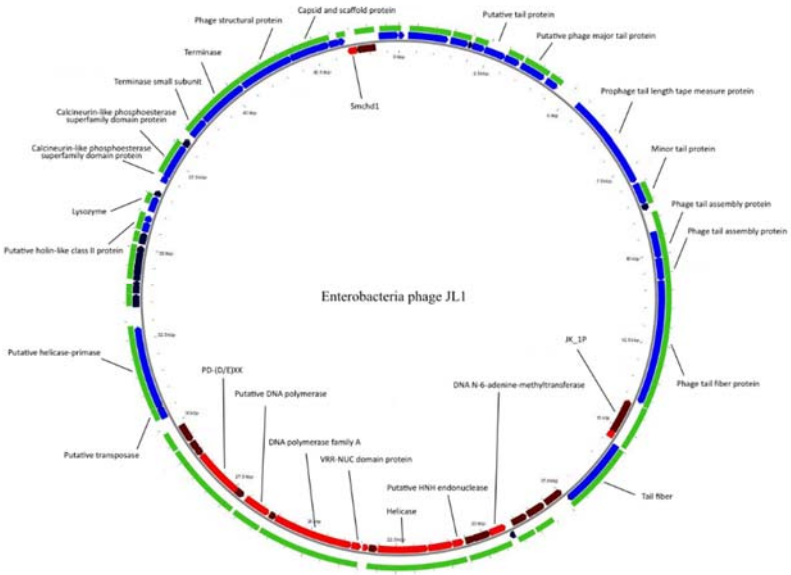


Figure 2. Genome map of phage SARL. A comparison was carried out and it was determined to be DW2.

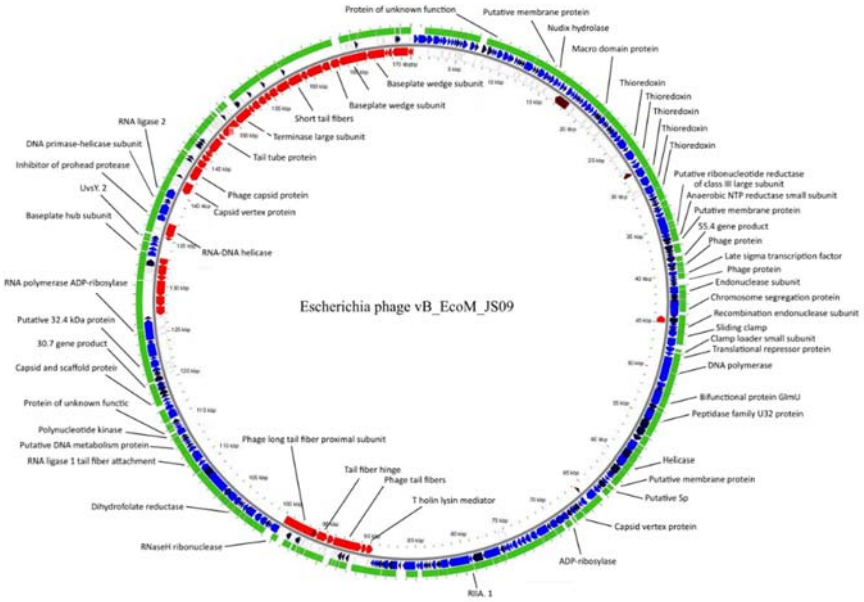
Table 3. Sensitivity of strains to SARL, DW2 and CS1 on the basis of efficiency of plaquing (EOP). (-) no clearing in spot assay. (+) clearing in spot assay but no plaques in plaque assay. EOP values are given for each phage. MRSA - Methicillin resistant *S. aureus.* MSSA - Methicillin sensitive *S. aureus.* hVISA - heterogenous vancomycin resistant *S. aureus.*

APC designation	Strain	Other name	Sensitivity designation	SARL	DW2	CS1
APC 317	*S. aureus*	DPC 5245	MSSA	1	1	1
APC 409	*S. aureus*	DPC 5247	MSSA	1.06	1.28	1.17
APC 2053	*S. aureus*	3594(II)ST36	MRSA	+	+	+
APC 993	*S. aureus*	ST290	MRSA	-	-	-
APC 994	*S. aureus*	ST291	MRSA	-	-	-
APC 995	*S. aureus*	ST295	MRSA	-	-	-
APC 996	*S. aureus*	ST299	MRSA	-	-	-
APC 1003	*S. aureus*	ST528	MRSA	-	-	-
APC 1004	*S. aureus*	ST530	MRSA	-	-	-
APC 1007	*S. aureus*	ST535	MRSA	-	-	-
APC 1009	*S. aureus*	ST544	MRSA	-	-	-
APC 2044	*S. aureus*	0.1345(II)ST5	MRSA	-	-	-
APC 2045	*S. aureus*	3596(IIV)ST8	MRSA	-	-	-
APC 2047	*S. aureus*	3488(VV)ST8	MRSA	-	-	-
APC 2048	*S. aureus*	3144(IIV)ST8	MRSA	-	-	-
APC 2049	*S. aureus*	E1038(IV)ST8	MRSA	-	-	-
APC 2050	*S. aureus*	E1185(IV)ST12	MRSA	-	-	-
APC 2051	*S. aureus*	E1174(IV)ST22	MRSA	-	-	-
APC 2052	*S. aureus*	0242(IV)ST30	MRSA	-	-	-
APC 2054	*S. aureus*	E1139(IV)ST45	MRSA	-	-	-
APC 2055	*S. aureus*	0.1239(III)ST239	MRSA	-	-	-
APC 2061	*S. aureus*	3581(1A)ST247	MRSA	-	-	-
APC 2062	*S. aureus*	0.1206(IV)ST250	MRSA	-	-	-
APC 1012	*S. aureus*	35197	hVISA	-	-	-
APC 1022	*S. aureus*	Newman	MSSA	-	-	-
APC 1024	*S. aureus*	RF122	MSSA	-	-	-
APC 79	*S. epidermidis*	DPC6010	n/a	-	-	-
APC 83	*S. hominis*	DPC6014	n/a	-	-	-

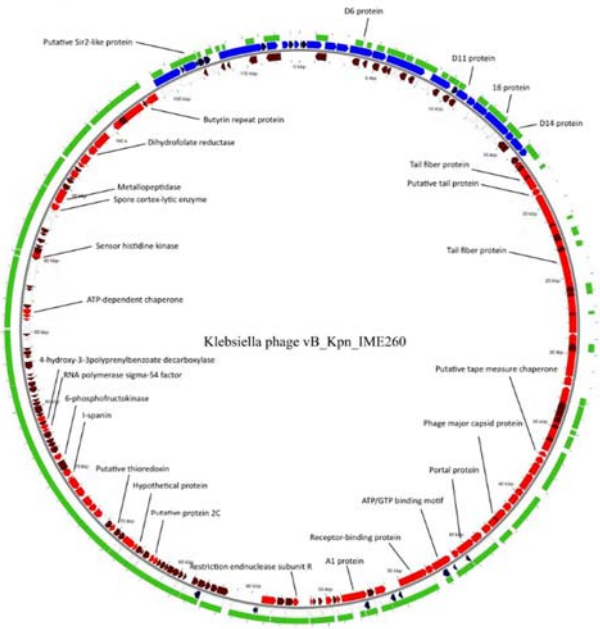
A.


Enterobacteria phage JL1
Capsid and scaffold protein
Phage structural protein
Terminase
Terminase small subunit
Calcineurin-like phosphoesterase superfamily domain protein
Calcineurin-like phosphoesterase superfamily domain protein
Lysozyme
Putative holin-like class II protein
Putative helicase-primase
PD-(D/E)XX
Putative DNA polymerase
Putative transposase
DNA polymerase family A
VRR-NUC domain protein
Helicase
Putative HNH endonuclease
DNA N-6-adenine-methyltransferase
JK_1P
Tail fiber
Phage tail fiber protein
Phage tail assembly protein
Phage tail assembly protein
Minor tail protein
Prophage tail length tape measure protein
Putative phage major tail protein
Putative tail protein
Smchd1

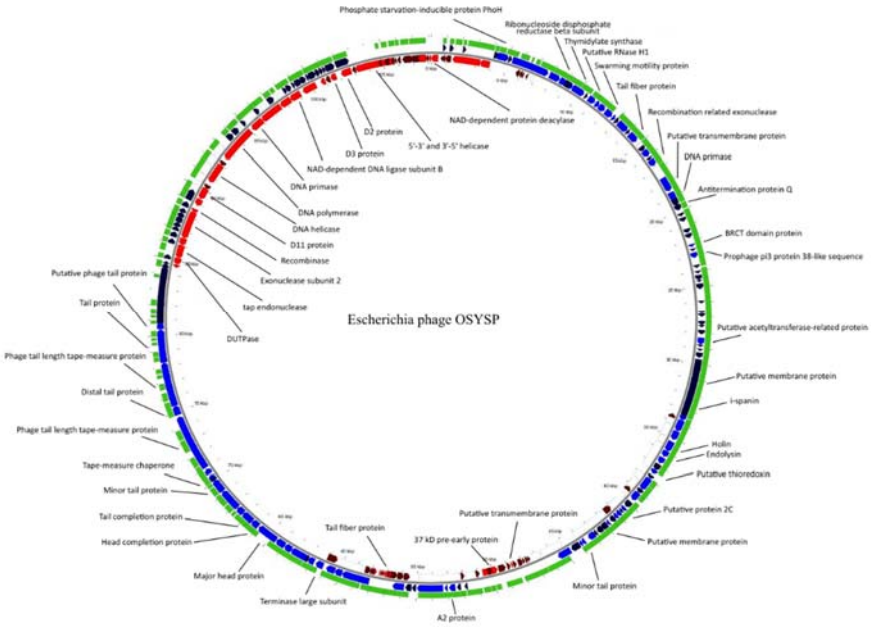
B.


Escherichia phage vB_EcoM_JS09
Protein of unknown function
Putative membrane protein
Nudix hydrolase
Macro domain protein
Thioredoxin
Thioredoxin
Thioredoxin
Thioredoxin
Thioredoxin
Putative ribonucleotide reductase of class III large subunit
Anaerobic NTP reductase small subunit
Putative membrane protein
55.4 gene product
Phage protein
Late sigma transcription factor
Phage protein
Endonuclease subunit
Chromosome segregation protein
Recombination endonuclease subunit
Sliding clamp
Clamp loader small subunit
Translational repressor protein
DNA polymerase
Bifunctional protein GlmU
Peptidase family U32 protein
Helicase
Putative membrane protein
Putative Sp
Capsid vertex protein
ADP-ribosylase
RIIA. 1
RNaseH ribonuclease
Dihydrofolate reductase
RNA ligase 1 tail fiber attachment
Putative DNA metabolism protein
Polynucleotide kinase
Protein of unknown functic
Capsid and scaffold protein
30.7 gene product
Putative 32.4 kDa protein
RNA polymerase ADP-ribosylase
Baseplate hub subunit
UvsY. 2
Inhibitor of prohead protease
DNA primase-helicase subunit
RNA ligase 2
RNA-DNA helicase
Capsid vertex protein
Phage capsid protein
Tail tube protein
Terminase large subunit
Short tail fibers
Baseplate wedge subunit
Baseplate wedge subunit
Phage long tail fiber proximal subunit
Tail fiber hinge
Phage tail fibers
T holin lysin mediator

C.


Klebsiella phage vB_Kpn_IME260
D6 protein
Putative Sir2-like protein
D11 protein
18 protein
D14 protein
Butyrin repeat protein
Dihydrofolate reductase
Tail fiber protein
Putative tail protein
Metallopeptidase
Spore cortex-lytic enzyme
Sensor histidine kinase
Tail fiber protein
ATP-dependent chaperone
4-hydroxy-3-3polyprenylbenzoate decarboxylase
RNA polymerase sigma-54 factor
6-phosphofructokinase
I-spanin
Putative tape measure chaperone
Phage major capsid protein
Portal protein
Putative thioredoxin
Hypothetical protein
Putative protein 2C
ATP/GTP binding motif
Receptor-binding protein
A1 protein
Restriction endnuclease subunit R

D.



E.

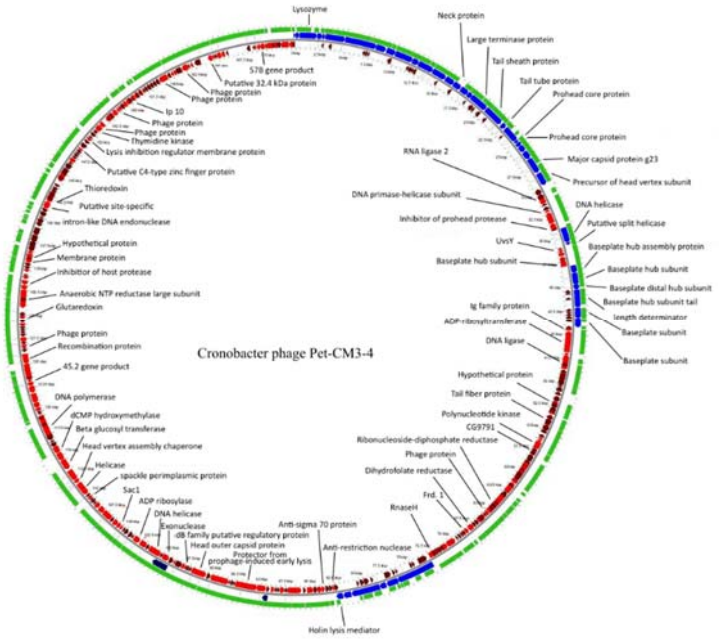


Figure 3. Genome map of phages isolated from environmental sources. A. Enterobacteria phage JL1 B. Escherichia phage vB_EcoM_JS09 C. Klebsiella phage vB_Kpn_IME260 D. Escherichia phage OSYSP E. Cronobacter phage Pet-CM3-4. Green lines represent pVOGs. Blue arrows represent genes on the forward strand. Red arrows represent genes on the reverse strand. Black arrows represent hypothetical proteins.

Table 4. Number of SNPs and changes in nucleotide broken down by group.

SNP	All groups	Intergenic regions	Non-synonymous changes	Synonymous changes
A→C	40	3	12	25
A→G	167	7	44	116
A→T	64	3	18	43
C→A	51	5	23	23
C→G	22	2	13	7
C→T	234	13	33	188
G→A	190	8	45	137
G→C	14	0	10	4
G→T	54	6	24	24
T→A	59	4	17	38
T→C	237	13	29	195
T→G	42	0	18	24
	1174	64	286	824

A.

B.

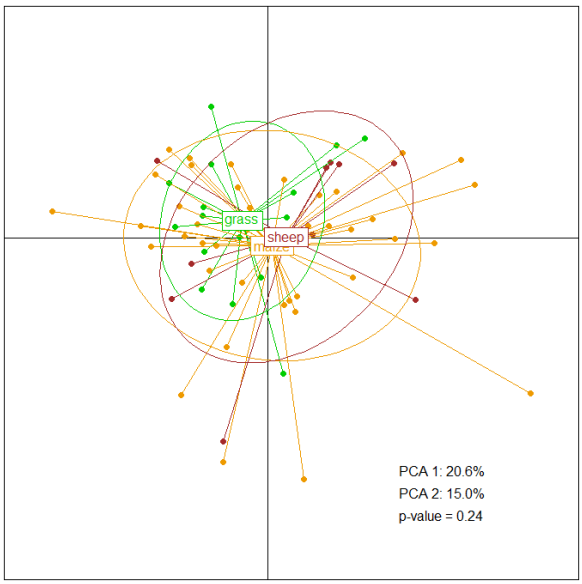


Figure 4. A. Box and whisker plot of number of SNPs of 70 isolated APCEc01 genomes compared to original APCEc01 genome from 2016. B. PCoA of pairwise distance between SNPs of 69 APCEc01 genomes (excluding A43M and original APCEc01 genome from 2016).

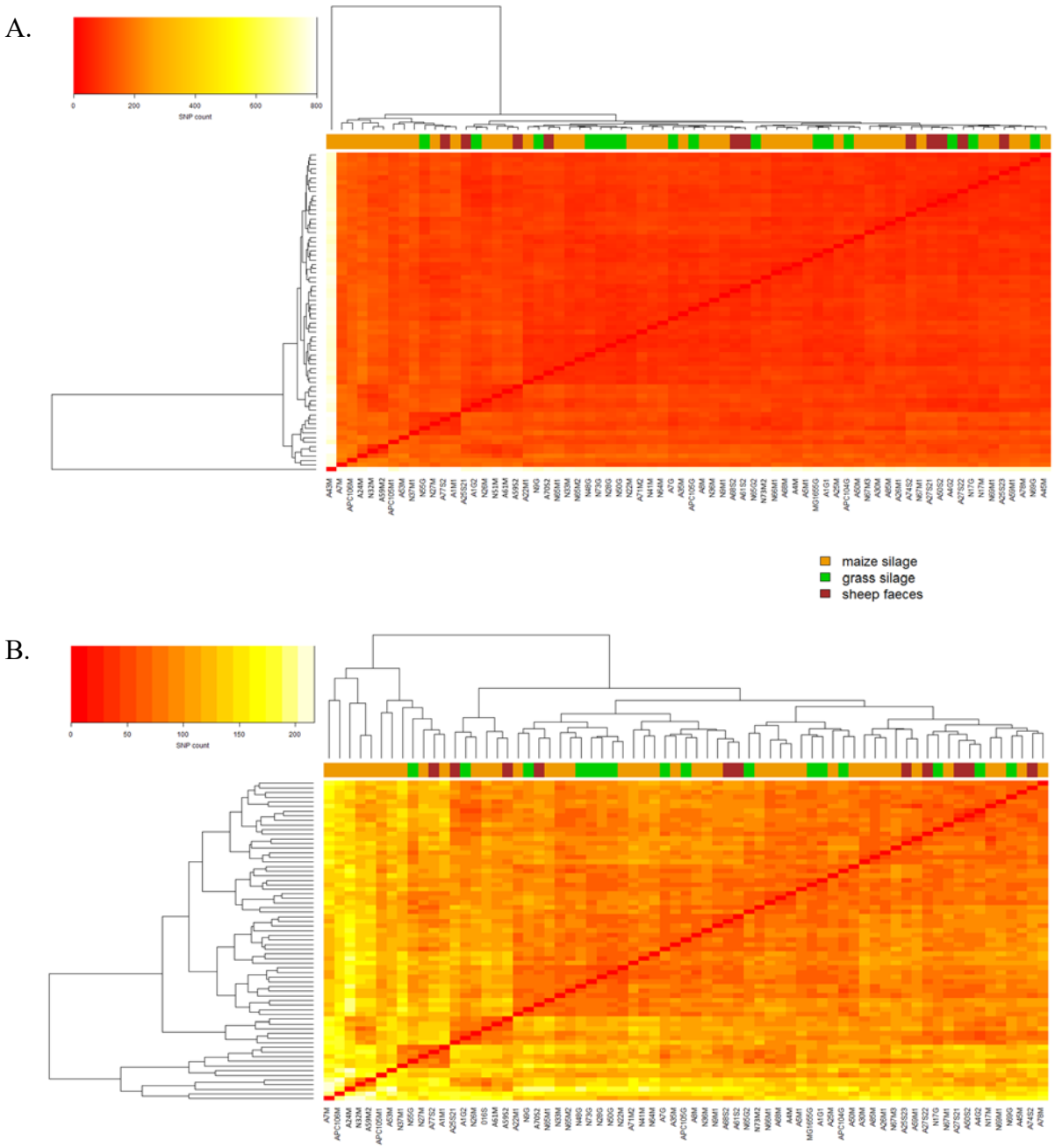


Figure 5. A. Heat map comparing SNP count between 70 APCEc01 genomes including A43M. B. Heat map comparing SNP count between 69 APCEc01 genomes excluding A43M.

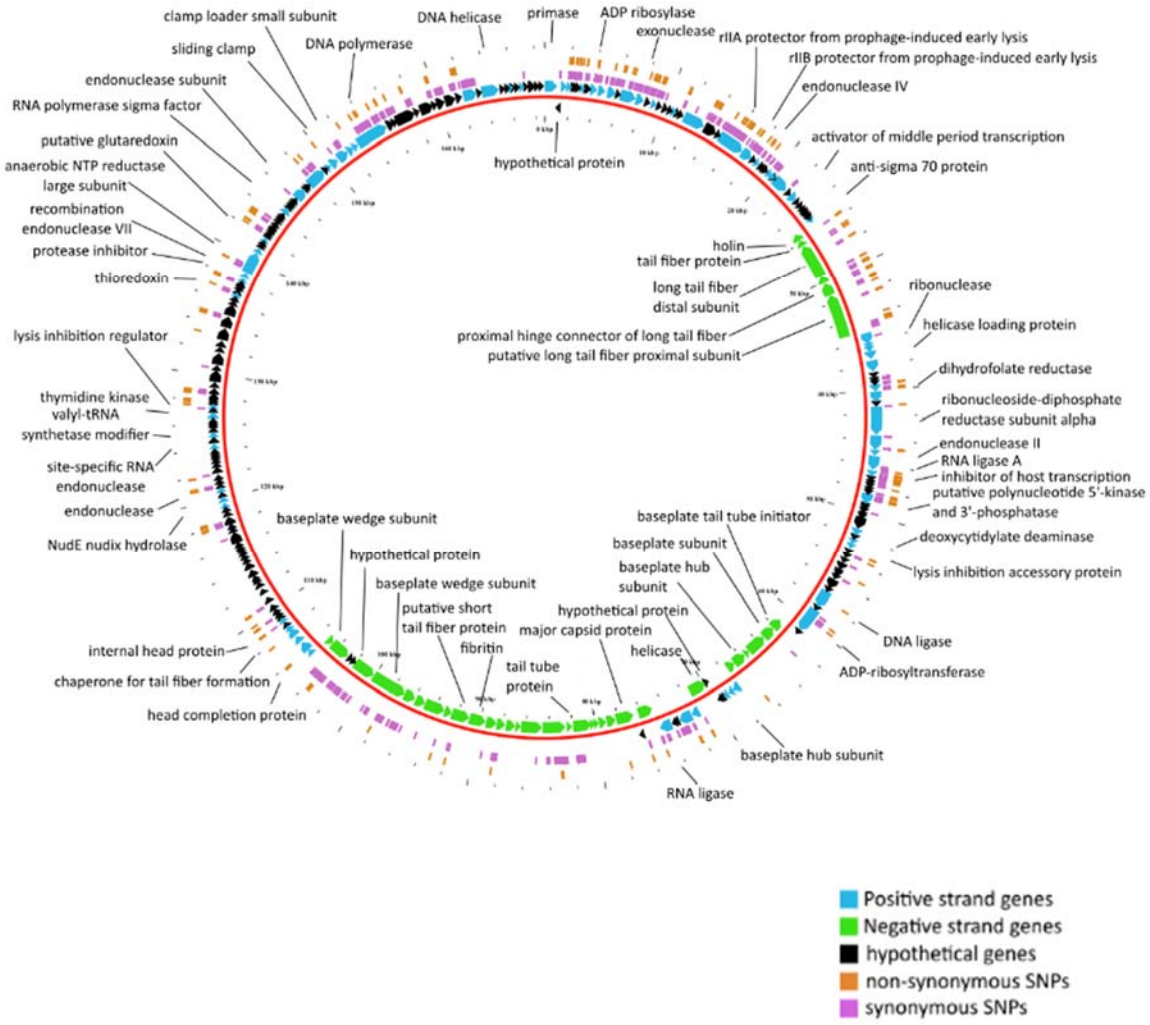


Figure 6. Synonymous and non-synonymous SNPs present in 70 APCEc01 genomes isolated from environmental samples compared to the original genome of APCEc01.

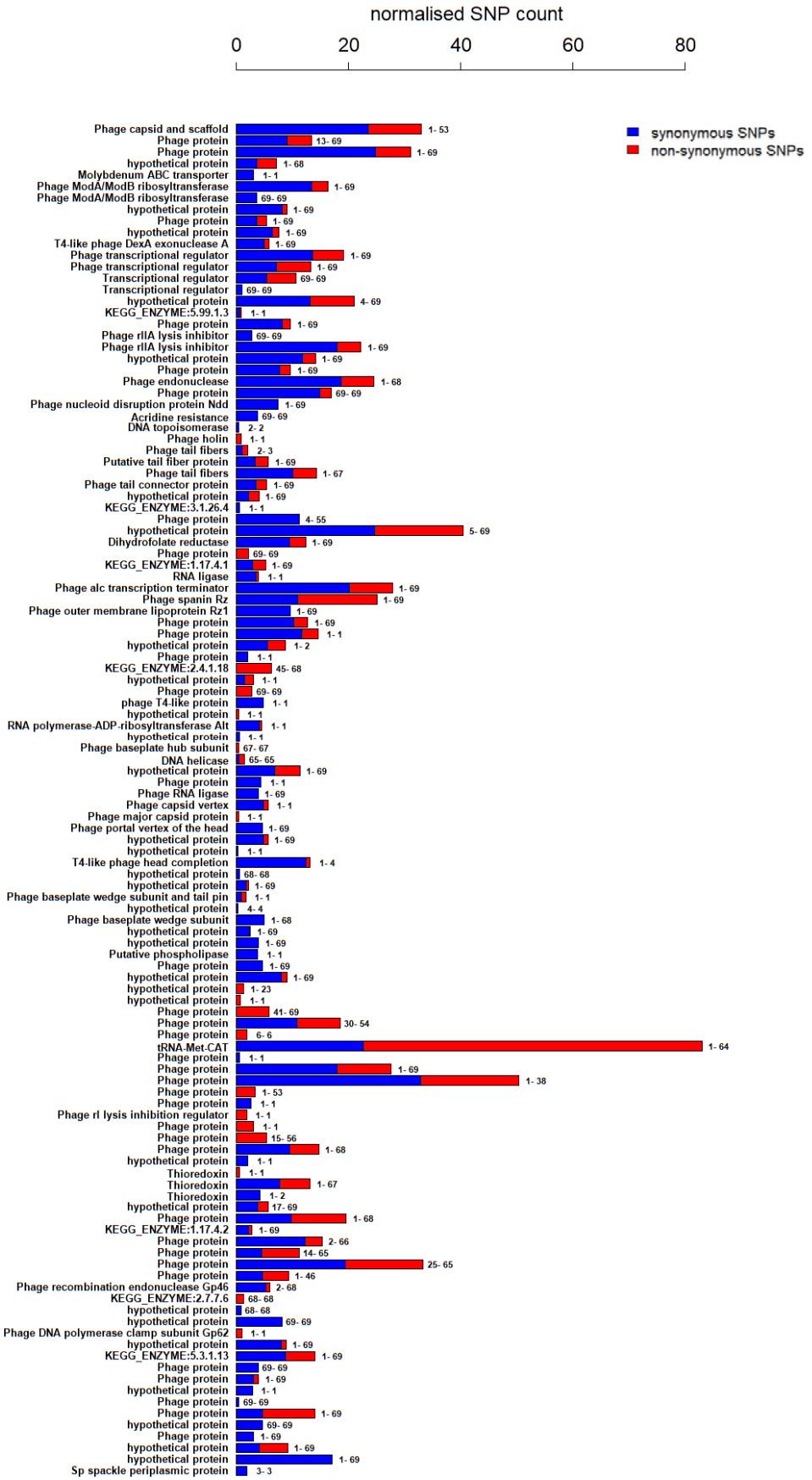


Figure 7. SNP distribution across all APCEc01 genes normalised by gene length.

Table 5. Cases in which two different SNPs were found at the same site and SNPs which introduced or involved stop codons.

Gene (position)	SNP position (+/- strand)	SNP	Original amino acid	SNP amino acid	Number of genomes with SNP
Phage rIIA lysis inhibitor (14883-17096)	16006 (+)	C→A	Serine	Stop codon	2
Phage endonuclease (18400-18876)	18612 (+)	A→G	Glycine	Glycine	1
Phage endonuclease (18400-18876)	18612 (+)	A→T	Glycine	Glycine	66
Phage endonuclease (18400-18876)	18613 (+)	A→T	Isoleucine	Phenylalanine	1
Phage endonuclease (18400-18876)	18613 (+)	A→C	Isoleucine	Leucine	66
Phage endonuclease (18400-18876)	18615 (+)	T→G	Isoleucine	Methionine	1
Phage endonuclease (18400-18876)	18615 (+)	T→A	Isoleucine	Isoleucine	66
Phage endonuclease (18400-18876)	18633 (+)	A→G	Threonine	Threonine	1
Phage endonuclease (18400-18876)	18633 (+)	A→C	Threonine	Threonine	68
Putative tail fiber protein (26377-29427)	27658 (-)	G→C	Tyrosine	Stop codon	67
Hypothetical protein (31300-35175)	34283 (-)	T→A	Aspartic acid	Valine	1
Hypothetical protein (31300-35175)	34283 (-)	T→C	Aspartic acid	Glycine	1
Phage protein (40967-41221)	41219 (+)	T→G	Stop codon	Glutamic acid	69
Hypothetical protein (53603-53971)	53678 (+)	C→T	Glutamine	Stop codon	1
Phage protein (117777-118463)	118117 (+)	T→A	Leucine	Stop codon	26
Phage protein (143308-143547)	143547 (+)	G→A	Stop codon	Stop codon	46

Chapter 5

Isolation of a novel jumbo bacteriophage effective against *Klebsiella aerogenes*

Published as:

Lewis R, Clooney A G, Stockdale S R, Buttimer C, Draper L A, Ross R P, Hill C. 2020. Isolation of a novel jumbo bacteriophage effective against *Klebsiella aerogenes*. Frontiers in Medicine 7:67.

Genome assembly and annotation was performed by Dr Adam G Clooney. Bioinformatic comparison of N1 and KCTC 2195 was performed by Dr Stephen R Stockdale.

Abstract

Increasing levels of bacterial resistance to many common and last resort antibiotics has increased interest in finding new treatments. The low rate of approval of new antibiotics has led to the search for new and alternative antimicrobial compounds. Bacteriophages (phages) are bacterial viruses found in almost every environment. Phage therapy was historically investigated to control bacterial infections and is still in use in Georgia and as a treatment of last resort. Phage therapy is increasingly recognised as an alternative antimicrobial treatment for antibiotic resistant pathogens. A novel lytic *Klebsiella aerogenes* phage N1M2 was isolated from maize silage. *Klebsiella aerogenes*, a member of the ESKAPE bacterial pathogens, is an important target for new antimicrobial therapies. *Klebsiella aerogenes* can form biofilms on medical devices which aids its environmental persistence and for this reason we tested the effect of phage N1M2 against biofilms. Phage N1M2 successfully removed a pre-formed *Klebsiella aerogenes* biofilm. Biofilm assays were also carried out with *Staphylococcus aureus* and Phage K. Phage K successfully removed a preformed *Staphylococcus aureus* biofilm. Phage N1M2 and Phage K in combination were significantly better at removing a mixed community biofilm of *Klebsiella aerogenes* and *Staphylococcus aureus* than either phage alone.

Introduction

Bacteriophages (phages) are viruses that require bacteria as a host for replication. Studies suggest that in many environments phages outnumber bacterial and archaeal cells 25:1 (1). Phages have been studied using culture based techniques since their independent discovery by Twort and D'Herelle in 1915 and 1917,

respectively. More recently, phages have been studied using culture independent methods such as the sequencing of environmental metagenomes which include bacteria, eukaryotic viruses and phages but also in studies of the virome alone (2). Interest has been growing in the role of phages in general human health (3) but also in specific conditions such as IBD (4). Despite the rise in viral metagenomics, cultivated phages are required for the development of interventions for use in phage therapy (5), food safety (6), and molecular biology (7).

The acronym ESKAPE has been applied to six bacterial pathogens that are of concern due to their ability to cause hospital acquired infections and the difficulty in treating them due to antibiotic resistance (8). The ESKAPE pathogens have shown increased incidence as causative agents in human disease and have increased antimicrobial resistance and negative outcomes. In 2017 the WHO compiled a global priority pathogens list prioritising the development of new antibiotics (9). The pathogens were chosen based on ten criteria; mortality, association with issues in healthcare settings and communities, the frequency of resistance, a 10-year trend of resistance, ease of transmission, how they could be prevented in hospital and community settings, and their current treatment. This list included critical priority targets carbapenem-resistant *Acinetobacter baumannii*, carbapenem-resistant *Pseudomonas aeruginosa* and carbapenem-resistant and 3rd generation cephalosporin-resistant Enterobacteriaceae (including *Klebsiella pneumoniae*, *Escherichia coli*, *Enterobacter* spp., *Serratia* spp., *Proteus* spp., and *Providencia* spp, *Morganella* spp). *Klebsiella aerogenes* was previously classified as *Enterobacter aerogenes* (10).

The rise in antibiotic resistance has led to a need for new antimicrobial compounds; however, the number of drug applications and approvals for antibiotics

has dropped significantly (11). For example, 19 new antibiotic applications were lodged in the early 1980s but only one was lodged between 2010 and 2012. This shortage of newly developed antibiotics has led to research into alternative antimicrobials (12). Western societies lost interest in phage therapy during World War II with the discovery of antibiotics but interest has been renewed due to the need for new antimicrobials. Phage therapy is currently used in Georgia against a range of pathogens, including *Shigella sonnei*, *Escherichia coli*, *Staphylococcus aureus* and *Pseudomonas aeruginosa*. Randomised controlled trials using phage therapy have been carried out such as PhagoBurn to treat burn wounds infected by *Pseudomonas aeruginosa* and the use of Biophage-PA to treat chronic otitis caused by antibiotic-resistant *P. aeruginosa* (13, 14). In the food industry, consumer distrust has increased towards chemical and artificial preservatives, leading to increased interest in alternative natural methods (15). Phages can also be used in food safety to reduce contamination from livestock during production, disinfect plants and equipment, reduce bacterial load on food at processing and during storage.

Much work has been carried out isolating phages from different sources. Faeces is a common source of phages, such as from ducks (16), cattle (17), pigs (18) and from sewage (19). Phages have been isolated from human sources, such as saliva (20), sputum samples, catheter tips and pleural effusions (21). Phages have also been isolated from environmental samples, such as seawater, from river estuaries (22), Arctic sea ice and melt ponds (23), soil (24), and yogurt and cheese factories (25).

The aim of this study was to isolate and characterise phages that could potentially be used for phage therapy or in food safety. A total of 163 bacterial strains were isolated from a healthy human faecal sample and used in conjunction with a bank of 79 known bacterial strains as prospective hosts for phages. These strains were

screened for phages against a human faecal filtrate and environmental samples collected from a nearby farm. Farm areas tested for phages were from cow slurry (n = 2), sheep faeces (n = 2), maize silage (n = 1) and grass silage (n = 1). A selection of phages were sequenced and analysed. One phage isolated on a *Klebsiella aerogenes* host, named phage N1M2, was chosen for further study as it had not been previously sequenced or described. Genomic and phylogenetic analysis revealed it to be a highly novel jumbo phage distantly related to *Pseudomonas* phage OBP, a phage previously described to possess homology to the PhiKZ-like jumbo phages (26).

Medical device-related infections are associated with increased morbidity, mortality and healthcare costs. Biofilms, which are commonly associated with medical devices, require treatment with antimicrobials in higher levels than planktonic cells (27). Phages have been investigated in human cases and mouse models of medical device infections. For example, phage coating significantly reduced the bacterial load and inflammation in a mouse model of MRSA orthopaedic implant infection (28). Moreover, phage significantly reduced bacterial numbers in a mouse model of implant infection caused by biofilm forming *P. aeruginosa* (29). Combinations of phage and antibiotics have also been successful in the treatment of a recurrent infection associated with a left ventricular assist device (30). In a case study, a combination of phage and antibiotics successfully treated a multidrug-resistant *P. aeruginosa* periprosthetic joint infection (31). The rationale behind the inclusion of phage was to aid in the breakdown of a biofilm matrix on the implant.

Urinary tract infections are a common nosocomial infection. *K. aerogenes* strains capable of biofilm production have been isolated from urinary catheters (32), and has also been found to be the causative agent in central venous catheter-related bloodstream infections (33). The effectiveness of phage N1M2 was therefore

investigated in biofilms. Mixed community infections are commonly associated with medical devices. Hola *et al*. found 465/534 urinary catheters were colonised by more than one bacterial strain (32). For this reason mixed community biofilms were also investigated. *Staphylococcus aureus* was chosen because it is one of the most commonly isolated bacteria from catheter-related urinary tract infections and bloodstream infections and methicillin-resistant *S. aureus* (MRSA) is one of the most feared hospital acquired infections (27). The antimicrobial potential of Phage K against *S. aurerus* is well documented (34). Phage K belongs to the family of *Herelleviridae* and falls within the genus of *Kayvirus* (35). Given the data to date more intensive research into the effectivity of phages in nosocomial infection is warranted especially given the current status of antimicrobial resistance to commonly used antibiotics.

Materials and methods

Isolation of bacterial hosts

Written consents were given according to study protocol APC055, approved by the Cork Research Ethics Committee (CREC). A healthy human faecal sample (1 g) was suspended in 5 ml of 1 X phosphate buffered saline (PBS) and 5 ml 40% glycerol and homogenised by vortexing for 5 min. Bacterial strains were isolated by diluting the faecal suspension in PBS and spread plating 100 µl of dilutions 10^{-3}-10^{-8} on MacConkey agar (Merck), Violet Red Bile agar (Merck), Brain Heart Infusion agar (Oxoid), MRS agar (Oxoid) and LB agar (Oxoid). Agar plates were incubated at 37˚C aerobically, and also under facultative anaerobic conditions using a gas jar with an Anaerocult A (Merck). Colonies of different morphology, size and colour were

isolated from the various different agar plates. Strains were streaked on their respective agar plates in triplicate to purify cultures for phage screening. A bank of known strains was also selected to be used for screening (Supp. Table 1). Bacterial strains were identified by 16S rRNA sequencing analysis using BLASTn. The 16S rRNA DNA was amplified by colony PCR using primers: F8-Fw: 5'-AGAGTTTGATCMTGGCTC-3' and R1509-Rv: 5'-GNTACCTTGTTACGACTT-3'.

Isolation and purification of phages

Human faecal and farm environmental samples (2 cow slurry, 2 sheep faeces, 1 maize silage and 1 grass silage) were collected and lysates were prepared immediately. Supernatants were prepared from samples using the following method. Sample (1g) was suspended in 10 ml of SM buffer (50 mM Tris-HCl; 100 mM NaCl; 8.5 mM $MgSO_4$; pH 7.5). Samples were homogenised by vortexing for 5 min, before centrifuging twice at 4,700X g for 10 min at 4˚C in a swing-bucket centrifuge to remove large particulates and bacterial cells. Supernatants were filtered twice through a 0.45 µm pore diameter filter. Bacterial strains were grown overnight in BHI at 37˚C with shaking (for aerobes) or without shaking (for facultative anaerobes). The overlay method was used for spot and plaque assays. BHI (1% agar w/v) was used as the base agar in a 100 mm X 15 mm petri dish. Plaque assays were performed by adding 400 µl of 1M $CaCl_2$ (final concentration 10mM), 100 µl of phage lysate and 100 µl of an overnight culture of bacterial host to 4 ml of soft BHI agar (0.5% agar w/v) kept at 50˚C. This mixture was poured on top of the BHI (1% agar w/v) base agar and allowed to solidify. Spot assays were performed by adding 400 µl of 1M $CaCl_2$ (final concentration 10mM) and 200 µl of an overnight culture of bacterial host to 4 ml of soft BHI agar (0.5% agar w/v) kept at 50˚C. This mixture was poured on top of the

BHI (1% agar w/v) base agar and allowed to solidify. 10 µl of phage sample lysates were pipetted on the agar and allowed to dry. Plates were incubated at 37°C for 24 to 48 h. Single plaque purification by propagation of a single plaque and plaque assay was carried out 3 times.

Viral DNA extraction, amplification, library preparation and sequencing

A 20 ml phage lysate with 4 ml of 2.5M NaCl and 50% polyethylene glycol (PEG) solution (final conc 0.4M NaCl and 8% (w/v) PEG) added was stored at 4˚C on ice overnight. Samples were centrifuged at 4700X g for 20 min at 4˚C in a swing bucket rotor. Supernatants were removed and pellets were dried for 5 min by inverting tube. Pellets were resuspended in 400 µl SM buffer (50 mM Tris-HCl; 100 mM NaCl; 8.5 mM $MgSO_4$; pH 7.5). 40 µl of 10X Nuclease Buffer (50 mM $CaCl_2$; 10 mM $MgCl_2$), was added and treated with 20 U of DNase I and 10 U of RNase I (final concentrations; Ambion) for 1 hr at 37˚C. Nucleases were inactivated at 70˚C for 10 min before samples were treated with 2 µl of freshly prepared 20 mg/µl Proteinase K for 20 min at 56˚C. Phage DNA extractions were performed using Norgen BioTek Corp Phage DNA Isolation Kit as described by the manufacturer starting at addition of Lysis Buffer B. The Elution Buffer (50 µl) was passed through the column twice to maximise DNA recovery. Viral DNA concentrations were equalised before paired-end Nextera XT library preparation (Illumina, San Diego, CA, USA) as described by the manufacturer. Metagenomic sequencing was performed using the Illumina MiSeq (Illumina Inc., San Diego, CA, USA) by generating 300 bp paired-end read libraries following the manufacturer's instructions.

Transmission electron microscopy

Phage lysate was concentrated by ultracentrifugation at 40,000 RPM for 2 hours at 4˚C. 180 ml of phage lysate was concentrated to 6 ml. The resulting phage preparation was placed on to a CsCl step gradient composed of 1.3, 1.5, and 1.7 g/ml layers and ultracentrifuged at 40,000 RPM for 3 hours at 4˚C. Resulting phage bands were collected and subjected to dialysis with two changes of SM buffer at 4˚C. 5 µl aliquots of the viral fraction were applied to Formvar/Carbon 200 Mesh, Cu grids (Electron Microscopy Sciences) with subsequent removal of excess sample by blotting. Grids were then negatively contrasted with 0.5% (w/v) uranyl acetate and examined at UCD Conway Imaging Core Facility (University College Dublin, Dublin, Ireland) by transmission electron microscope.

Bacterial DNA extraction, library preparation and sequencing

N1 DNA was extracted from an overnight culture using a Qiagen DNeasy Blood and Tissue kit as per the manufacturer's instructions with pre-treatment for Gram-negative bacteria. N1 DNA was sequenced using the Oxford Nanopore MinION as per manufacturer's instructions.

Phage propagation and plaque assays

After isolation of phage N1M2 propagation was carried out in Tryptic Soy Broth (TSB) with calcium boroglucinate (final concentration 10mM). Plaque assays were carried using 1% Tryptic Soy Agar (TSA) base agar and 0.4% TSA overlay with calcium boroglucinate (final concentration 10mM). Phage N1M2 host range was established on 1.5% LB base agar and a 0.2% LB agarose overlay with $CaCl_2$ (final concentration 10mM) by carrying out plaque assays against a range of bacteria. The efficiency of plaquing was calculated by dividing the titre of phage N1M2 on the strain

to be tested by the titre of phage N1M2 on strain N1. For host range assay bacteria were grown overnight in LB broth at 37˚C under aerobic conditions. Phage K propagation was carried out in TSB with calcium boroglucinate (final concentration 10mM). Phage K plaque assays were carried using 1% TSA base agar and 0.4% TSA overlay with calcium boroglucinate (final concentration 10mM).

N1 Bioinformatic analysis

K. aerogenes N1 draft genome and *K. aerogenes* KCTC 2190 genome were compared using Easyfig v2.2.2 (57). Genomes were compared using BLASTn with default settings except for minimum length and minimum identity which were changed to 100 bp and 50% respectively. PHASTER was used to predict possible prophage regions in N1 and KCTC 2190 (58). JGI IMG/M genome browser was used to analyse function of KTCT 2190 genes (59).

Phage N1M2 Bioinformatic analysis

Sequencing reads of phages were quality filtered and assembled into contigs using SPAdes meta (60). BLASTn NT database was used to ensure the phage N1M2 sequence was not similar to that of a known phage (61). Phage N1M2 genome was annotated using VIGA (https://github.com/EGTortuero/viga; (62)). Functional annotation of ORF gene products and amino acid identity was established using BLASTn and InterProScan (63). Transmembrane helices in proteins were predicted using TMHMM (64). LipoP (65) was used for predictions of lipoproteins. The molecular weights and isoelectric points of the predicted ORFs were determined using https://web.expasy.org/compute_pi/. Amino acid and codon usage statistics were determined using the University of Georgia online tool (http://www.cmbl.uga.edu/software/codon_usage.html). The presence of transfer

RNA genes was determined using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/; (66) and ARAGORN (http://130.235.46.10/ARAGORN/; (67)). BLASTn search (evalue cut-off 1e-10) was performed against NCBI nt database and bacterial section of NCBI RefSeq genomes database release 89 to align to bacterial tRNAs. Potential Rho-independent terminators were identified using ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold; (68)) and then confirmed with MfoldQuikFold (http://unafold.rna.albany.edu/?q¼DINAMelt/Quickfold; (69)) using RNA energy rules 3.0. Potential promoters were identified by extracting sequences 100 bp upstream of each ORF using FeatureExtract 1.2L (light) Server (http://www.cbs.dtu.dk/services/FeatureExtract/; (70)) and submitting these sequences to MEME (Multiple Em for Motif Elicitation) (http://meme-suite.org/tools/meme; (71)). MEME discovers novel, ungapped motifs of recurring and fixed-length patterns between submitted sequences. In this case it looked for similar motifs in the sequences submitted, 100 bp upstream of each ORF, where promoters would be found. The annotated genome was visualised using GView (72). Multiple sequence alignment was carried out with MEGA7 (73) using MUSCLE (74). Phylograms were then constructed using the major capsid protein, the large terminase and DNA polymerase being used among phage N1M2 and other jumbo phages. The evolutionary history of phage N1M2 was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (75) in MEGA 7. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analysed (76). The whole genome phylogenetic trees based on nucleotides and amino acid sequences were generated by Victor (77) using the Genome-BLAST Distance Phylogeny (GBDP) (78) method under settings recommended for prokaryotic viruses. Bacterial CRISPR spacer database (made in

house using PILER-CR v1.06 from NCBI RefSeq genomes database, release 89) was used as a query for BLASTn search (evalue cut-off 1e-5) against the phage genome.

Biofilm assays

96-well plates were filled as follows based on Dalmasso, *et al.* (20). Each well was filled with 200 µl TSB with or without 1% glucose inoculated at 1% with an overnight culture of the relevant strain. For mixed biofilms 1% of an overnight culture of N1 and 1% of an overnight culture of DPC 5247 (Methicillin sensitive) was added. Plates were incubated at 37˚C for 48 h to allow the biofilm to form. Broth containing planktonic cells was carefully removed to avoid disruption of the formed biofilm. 100 µl TSB containing 20mM calcium boroglucinate and 100 µl of phage dilution was added to each well. SM buffer was added to control wells. Plates were incubated at 37˚C for 48 hr as indicated. 24 h biofilm formation was also tested as was phage addition for 72 hr. For investigating stopping the formation of a biofilm 50 µl phage was added at the same time as the 200 µl overnight culture of the relevant strain. After incubation, broth containing planktonic cells was carefully removed. Wells were washed twice with 150 µl phosphate buffered saline (PBS). An XTT/menadione assay was carried out as follows. XTT/menadione solution was prepared by adding 0.01 g XTT to filter sterilised water and filter sterilising using a 0.22 µm filter. 0.027 g menadione was added to 10ml acetone. 10 µl of menadione acetone solution was added to 20 ml of XTT solution. 100 µl of XTT/menadione solution was added to each well of the 96-well plate and incubated at 37˚C in the dark for 2 hours. The absorbance was then measured at a wavelength of 492nm in a plate reader. The XTT/menadione assay relies on the reduction of tetrazolium salt 2,3-bis[2-methyloxy-4-nitro-5-sulfophenyl]-2H-tetrazolium-5-carboxanilide (XTT) by metabolically active cells to

an orange/yellow water-soluble formazan derivative that can be quantified colorimetrically meaning only live cells are counted (79).

N1 and DPC 5247 numbers were quantified in a mixed biofilm. The experiment was carried out once but counts were carried out in duplicate. Biofilms were formed and treated as previously described. After incubation, broth containing planktonic cells was carefully removed. Wells were washed twice with 150 µl PBS. 200 µl PBS was added to each well and mixed by pipetting up and down. All the wells in a row were combined and serially diluted 1:10 in PBS. 100 µl of the relevant dilution was spread plated on UTI ChromoSelect agar. Plates were incubated at 37°C for 24 hours.

Statistical analysis

For XTT assays 2 rows per 96-well plate were used per condition to be tested and all experiments were independently performed 3 times. For enumeration of N1 and DPC 5247 the experiment was carried out once but counts were carried out in duplicate. Graphs were prepared using GraphPad Prism and are presented as mean values of a single experiment that represent the trend seen in the triplicate experiment. Error bars in the figures indicate standard error of the mean. 1 way ANOVA was used to determine the significance of differences between controls and treated samples.

Results

Genome information of strain N1

Bacterial strain N1 was isolated from a healthy human faecal sample using MacConkey agar and identified as *Klebsiella aerogenes* by 16S rRNA sequencing. It

was then genome sequenced using the Oxford Nanopore MinION. This was classed as a draft genome due to the high error rate associated with MinION sequencing (36). Strain N1 has a chromosome of 5167877 bp with a GC content of 55.3%. The chromosome encodes 14509 predicted open reading frames (ORFs) ranging in size from 90 to 3315 bp. N1 contained two plasmids of 72172 bp and 51484 bp with GC contents of 43.6% and 47.2% respectively. Plasmid 1 encoded 158 ORFs ranging in size from 93 to 636 bp. Plasmid 2 encoded 123 ORFs ranging in size from 114 to 660 bp. No resistance genes or tRNA were predicted in plasmids 1 and 2. Twenty-four tRNA genes were predicted in the chromosome (Table 1). N1 was compared to the reference genome *Klebsiella aerogenes* KCTC 2190. KCTC 2190 and N1 are similar in gene order with no inversions or large areas of difference (Fig. 1A). Two intact prophage regions were predicted in KCTC 2190 while one intact, two questionable, and two incomplete prophage regions were predicted in N1 (Fig. 1B). The Pfam categories of KCTC 2190 encoded functions were used as indicators of the functions associated with N1 due to their similarity (Fig. 1C). Almost 12% of KCTC 2190 genes were of unknown function which is a common occurrence. KCTC 2190 included a number of antibiotic resistance genes and a number of different antibiotic resistance genes were predicted in the N1 genome confirming its importance in the rise of antibiotic resistance and as an ESKAPE pathogen (Supp. Table 2). Genes encoding MarR, EmrR, and AcrA are present in N1 and all have been linked to multidrug resistance in Enterobacteriacae (37). AcrA has been associated with resistance to carbapenem antibiotics. A number of efflux pumps were present in N1 and these are a common mechanism of antibiotic resistance in *Klebsiella aerogenes*.

General information and characterisation of phage N1M2

Phage N1M2 was isolated against *Klebsiella aerogenes* N1 from maize silage. Phage N1M2 had a genome size of 253367 bp, placing it among the double stranded DNA jumbo phages. The mean GC content is 40.9%. EM showed phage N1M2 to be a *Myoviridae* phage as evidenced by the long contractile tail (Fig. 2A). Broken tail and head structures were visible. The contraction of the tail sheath made the tail core visible (Fig. 2B). No tail fibers were visible in the EM images. The average capsid size was 113 nm (± 6 nm) X 101 nm (± 7 nm) (N=7). The average tail size was 158 nm (± 11 nm) X 21 nm (± 1 nm) (N=7).

The host range of phage N1M2 was tested against a range of bacterial strains (Table 3). In addition to *Klebsiella aerogenes* phage N1M2 also infected *Klebsiella aerogenes* NCIMB 10102 and *Klebsiella pneumoniae* NCIMB 13218. Plaque morphology was similar on all three strains but the efficiency of plaquing was much lower for Klebsiella pneumoniae NCIMB 13218 compared to N1 and *Klebsiella aerogenes* NCIMB 10102 (0.017 ± 0.009 compared to 1 ± 0 and 0.667 ± 0.17, respectively).

The genome of phage N1M2

The phage N1M2 genome is comprised of 257 ORFs of which 39 are encoded on the minus strand (Supp. Table 3). The putative gene products ranged in size from 29 to 3375 amino acids. All ORFs were predicted to start with AUG. Based on a combination of BLAST and InterProScan 85 putative gene products were identified. Of the 172 unidentified putative gene products 106 aligned to unnamed protein products in *Pseudomonas* phage OBP. Of the unnamed protein products of *Pseudomonas* phage OBP the identity ranged from 21-88%. Three tRNA genes were

predicted and all were contained within a 947 bp region located between 249035 bp and 249982 bp (Table 1). The tRNA genes were similar to those found in *Escherichia* species, *Salmonella enterica*, *Shigella* species and Stx2-converting phage, and an *Enterobacteria* phage. No integrase, excisionase or repressor genes were identified, suggesting that the phage follows a lytic lifestyle (Supp. Table 3).

The potential promoter TBYAWWWWWTTTCARRYAKATATTATYWAAGTGWA was identified at 31 locations in the genome (Supp. Table 4). The promoter was compared to similar promoters in jumbo phages related to phage N1M2 (Table 2). Fifty-six potential rho-independent terminators were predicted throughout the genome of phage N1M2 (Supp. Table 5).

At the DNA level phage N1M2 was found to have little homology to phage genomes available on public databases, with the closest match being the aforementioned *Pseudomonas* phage OBP (BLASTn analysis coverage: 55%, identity: 72.73%). The majority of proteins which aligned to known proteins on BLAST aligned to this phage (Supp. Table 3) (Fig. 3B). To investigate the relationship of phage N1M2 to other phages phylogenetic trees were constructed using the nucleotide and amino acid whole genome sequences (Fig. 4). To further characterise the relationship between phage N1M2 and some of these phages phylogenetic trees were constructed using the amino acid sequences of the major capsid protein, large terminase and DNA polymerase (Fig. 5).

Single strain biofilms using strain N1

Biofilms of *K. aerogenes* N1 were formed under various conditions and XTT assays were used to assess the efficiency of phage N1M2 in destroying existing

biofilms and preventing biofilm formation (Fig. 6). Biofilm experiments were performed three times with duplicate readings and the data from one experiment is presented as a representative of the triplicate results. Phage N1M2 at an MOI of 10 or 100 for 48 hours reduced an N1 biofilm formed in the presence of glucose over 48 hours but this did not achieve significance ($P>0.05$) (Fig. 6A). Phage K at an MOI of 10 and 100 also had no significant effect on the N1 biofilm (Fig. 6A) which was expected since phage K does not infect strain N1. Similarly, phage N1M2 at an MOI of 10 or 100 for 48 hours did not significantly reduce an N1 biofilm formed without glucose over 48 hours ($P>0.05$) (Fig. 6B). However, phage N1M2 applied at an MOI of 10 or 100 for 72 hours significantly reduced an N1 biofilm formed over 48 hours without glucose ($P<0.001$ and $P<0.05$ respectively) (Fig. 6C). The reduction by phage N1M2 at an MOI of 10 or 100 was not significantly different. Phage K at an MOI of 10 applied for 72 hours did not reduce the N1 biofilm (Fig. 6C). In an N1 biofilm formed over 24 hours without glucose phage N1M2 applied at an MOI of 10 for 48 hours once again significantly reduced the biofilm ($P<0.001$) (Fig. 6D). Once again Phage K had no effect.

As well as destroying preformed biofilms the ability of phage N1M2 to impede the formation of biofilms was investigated. To do this phage N1M2 was added when the bacteria was inoculated. Phage N1M2 applied at an MOI of 10 did not prevent a biofilm from forming or reduce the level of biofilm formed. This was seen in biofilms formed with (Fig. 6E) and without glucose (Fig. 6F).

Single strain biofilms using DPC 5247

Biofilms of *S. aureus* DPC 5247 were formed under various conditions and the impact of Phage K was measured using XTT assays (Fig. 7). Phage K at an MOI of

10 or 100 applied for 48 hours significantly reduced a DPC 5247 biofilm formed in the presence of glucose over 48 hours ($P<0.001$) (Fig. 7A). Phage N1M2 at an MOI of 10 had no effect on the DPC 5247 biofilm (Fig. 7A) which was expected since phage N1M2 cannot infect DPC 5247. The reduction due to Phage K at an MOI of 10 was significantly better than that of Phage K at an MOI of 100 ($P<0.001$). Phage K at an MOI of 10 applied for 48 hours also significantly reduced a DPC 5247 biofilm formed without glucose over 48 hours ($P<0.001$) (Fig. 7B). Phage N1M2 at an MOI of 10 applied for 48 hours had no significant effect on a DPC 5247 biofilm formed without glucose over 48 hours.

Mixed community biofilms using N1 and DPC 5247

Mixed community biofilms of N1 and DPC 5247 were formed and treated with phage N1M2 and Phage K alone and in combination. XTT assays were carried out to assess their effect (Fig. 7). Phage N1M2 and Phage K at MOIs of 10, alone and in combination, applied for 48 hours did not significantly reduce a mixed biofilm of N1 and DPC 5247 formed in the presence of glucose over 48 hours ($P<0.001$) (Fig. 7C). Equally, in a mixed community biofilm of N1 and DPC 5247 grown for 48 hours without glucose phage N1M2 alone and Phage K alone at MOIs of 10 applied for 48 hours did not significantly reduce the biofilm. However, phage N1M2 and Phage K at MOIs of 10 in combination significantly reduced the biofilm ($P<0.001$) (Fig. 7D).

Untreated and phage-treated biofilms were quantified by dilution and spread plating on UTI ChromoSelect agar (Supp. Figure 1). N1M2 and Phage K at MOIs of 10, alone and in combination, did not reduce the N1 portion of the community. However, Phage K at an MOI of 10 and N1M2 and Phage K at MOIs of 10 in combination reduced the DPC 5247 members of the community.

Discussion

Phage N1M2 is the first jumbo phage identified which kills *Klebsiella aerogenes*. Jumbo phages have a number of common properties which differentiate them from smaller phages (38). Jumbo phages are tailed phages with genomes larger than 200000 bp with large virions. The genes are distributed throughout the genome rather than in organised blocks as is typical of smaller phage genomes. Also jumbo phages contain more genes associated with biological processes such as genome replication, nucleotide metabolism, and lysis of host cell peptidoglycan than smaller phages. Some jumbo phages have more than one DNA polymerase or RNA polymerase gene. Jumbo phages are phylogenetically unrelated to smaller phages. This was evident in our analysis as phage N1M2 showed no similarity to small *K. aerogenes* phages and was most closely related, albeit only distantly, to other jumbo phages (Fig. 3). Jumbo phages are isolated more rarely than smaller phages for a number of reasons related to their large size. Jumbo phages often cannot diffuse in semi-solid media meaning they are not always visible as plaques. Their large size also means they can be removed when filtering out bacteria as they cannot pass through smaller filter pores.

The genome of phage N1M2 was suggested to be circularly permuted based on its similarity to *Pseudomonas* phage OBP, its closest relative (26). *Pseudomonas* phage EL, the closest relative of *Pseudomonas* phage OBP, is also circular permuted. When aligned using Mauve, phage N1M2 and OBP showed two areas of similarity, a short region and a long region, although within these similar regions are a number of regions with no homology (Fig. 3B). The areas of phage N1M2 that align to OBP ranged from 21-88% identity. In the whole genome trees at nucleotide and protein level phage N1M2 and *Pseudomonas* phage OBP formed a separate, strongly

bootstrapped clade (Fig. 4). The major capsid protein, large terminase and DNA polymerase trees supported these whole genome trees (Fig. 5).

A potential promoter TBYAWWWWWTTTCARRYAKATATTATYWAA-GTGWA was identified at thirty-one locations in the phage N1M2 genome (Supp. Table 4). A possible -10 sequence, KATA, was present but unclear. MEME uses K to signify that G or T could be present in a position of the motif identified. No conserved -35 sequence was present. No known sigma factors were present in the N1M2 genome (Supp. Table 3). Similar sequences have been identified in OBP and *Pseudomonas* phage 201phi2-1 (26), *Pseudomonas* phage EL (39) and *Pseudomonas* phage phiKZ (40). The promoter found in phage N1M2 was most similar to that of OBP but also similar to the TATATTAC block found in phiKZ (Table 2). The OBP promoter was suggested as phage specific based on its lack of similarity to sequenced *Pseudomonas* genomes and was deemed to be an early promoter as they were located at the start of operons associated with early genes but not middle or late genes (26). The potential promoter was situated before ORFs with predicted products with unknown functions but also with predicted products such as RNA polymerase beta subunit, putative virion structural protein, and putative lytic murein transglycosylase.

Despite the clinical importance of *K. aerogenes* little work has been carried out isolating and characterising *K. aerogenes* phages. There are only four phage genomes publically available: vB_EaeM_φEap-1 (NC_028772), F20 (JN672684) (41), vB_EaeM_φEap-2 (NC_028695) (42) and vB_EaeM_φEap-3 (KT321315) (43). K. Verthe*, et al.* (44) also isolated UZ1 *K. aerogenes* phage but no genome sequence is available. None of these phages are jumbo phages, ranging from 39133 bp to 175814 bp. The host range of phage N1M2 was tested against a range of bacterial strains selected based on these publications. It is difficult to establish the host range of *K.*

aerogenes phages in general based on this small amount of data. So far the host range of *K. aerogenes* phages appears to vary. In some cases it is limited to a single *K. aerogenes* strain, although only a small number of strains were tested (44). In other cases it can include multiple *K. aerogenes* strains but no other genera or species (42, 43). Phage N1M2 infected the two *K. aerogenes* strains tested and also *Klebsiella pneumoniae*, although the efficiency of plaquing of phage N1M2 was much lower on *Klebsiella pneumoniae* 13218 (Table 3). Infection outside of the genus has not been previously seen in *K. aerogenes* phages. *Klebsiella pneumoniae* is a leading cause of hospital acquired infections and has been linked to a number of illnesses such as wound infections, soft tissue infections, urinary tract infections, pyogenic liver abscess, pneumonia, meningitis and neonatal sepsis (45). The emergence of multi drug-resistant *Klebsiella pneumoniae* has made it of critical concern for human health and the fight against antibiotic resistance. Phages have been investigated for the treatment of experimentally induced *Klebsiella pneumoniae* burn wound infections and pneumonia (46, 47). Phages have also been used to successfully treat a multi-drug resistant *Klebsiella pneumoniae* wound infection (48). Phage N1M2 showed no homology to known jumbo phages *Klebsiella* phage K64-1 and vB_KleM-RaK2 which both infect *Klebsiella* strains.

It is possible that the increased host range of phage N1M2 is related to the presence of tRNA along with the increased number of genes for genome replication, nucleotide metabolism and host cell-wall lysis (38). For example the entire genome of *Pseudomonas aeruginosa* Phage KZ can be transcribed using its own RNA polymerase with no significant contribution from the host RNA polymerase (49). tRNAs in phages correspond to codons that are more commonly used in the phage than their bacterial host, especially in genes encoding structural proteins. This reduces

the phages reliance on the host cell machinery for replication and increases host range (50). The presence of tRNAs for codons that are preferentially expressed in phage N1M2 compared to N1 was apparent in two out of three cases in phage N1M2 (Table 1). Phage N1M2 has three tRNAs for the codons CAT, TCT and GTT. The codon usage of CAT (for Histidine) by phage N1M2 is less than that of N1, 1.04% and 1.23% respectively, with a ratio of phage/host of 0.85. Phage N1M2 uses codon TCT (for Serine) at a frequency of 2.13% while N1 uses TCT at a frequency of 0.73% with a ratio of phage/host of 2.92. Phage N1M2 uses codon GTT (for Valine) at a frequency of 3.40% while N1 uses GTT at a frequency of 1.43% with a phage/host ratio of 2.38.

The jumbo phages also have a number of genes involved in nucleotide metabolism such as thymidylate synthase, thymidylate kinase, ribonucleoside diphosphate reductase, NrdB, and dihydrofolate reductase and multiple genes encoding β and divided β' subunits specific for multisubunit bacterial RNA polymerases and virion associated RNA polymerases (26). These genes can aid or replace the function of host genes. Of these phage N1M2 contained genes for thymidylate synthase (ORF 184), thymidylate kinase (ORF 249 but of low homology), ribonucleotide diphosphate reductase beta subunit NrdB (166), and dihydrofolate reductase (ORF189). Phage N1M2 contained three genes for RNA polymerase β subunits (ORF37, ORF51 and ORF52) and three putative virion-associated RNA polymerase beta subunits (ORF64, ORF222 and ORF223). Phage N1M2 also contained genes for thymidine kinase (ORF196). A putative electron carrier glutaredoxin (ORF253) was present which could support NrdB function. Again all of these genes showed similarity to OBP.

Phage N1M2 applied for 48 hours had no significant effect on a 48 hour preformed biofilm (Fig. 6A). Phage N1M2 could also not stop a biofilm from forming

over 48 hours (Fig. 6E). However, phage N1M2 applied for 72 hours reduced a biofilm formed over 48 hours and phage N1M2 applied for 48 hours reduced a biofilm formed over 24 hours (Fig. 6C and Fig. 6D). The importance of allowing an adequate time for phage to have an effect can be seen. Hosseinidoust, *et al.* found that after the addition of phage biofilms increased in some cases up to five times that of untreated controls after 24 hours but phage treated biofilms at 48 hours and 72 hours were similar to untreated controls (51). They also found that phage did not significantly reduce biofilms until 48 or 72 hours after phage application, before this point the phage treated biofilm was equal to or greater than an untreated biofilm. In some circumstances phage treatment can cause an increase in viable cell counts of biofilms. González, *et al.* found that phage treatment, using a phage that did not infect *E. faecium* or *L. plantarum*, of mixed biofilms formed over 5 hours significantly increased viable cell counts of *E. faecium* and *L. plantarum* while in 24 hour biofilms *L. plantarum* was slightly decreased and *E. faecium* was similar to untreated controls (52). Mixed species biofilms are often more resistant to antibiotics than single species biofilms.

We tested 48 hours of phage application to a 48 hour mixed species biofilm to determine if the inclusion of another phage could improve the activity of phage N1M2 in a situation where it had previously been ineffective. Phage K alone could effectively reduce *S. aureus* DPC 5237 in a single species biofilm (Fig. 7B). The reduction by Phage K of biofilms has been previously seen (53, 54). Phage N1M2 alone could not reduce N1 numbers in a mixed biofilm. This was expected as phage N1M2 alone could not reduce N1 numbers in a single species biofilm. In XTT assays a combination of phage N1M2 and Phage K was more effective than Phage K alone (Fig. 7D). 48 hours of phage application to a 48 hour biofilm was quantified to see if the inclusion of another phage could improve the activity of N1M2 in a situation where it had

previously been ineffective (Supp. Figure 1). Phage K alone could effectively reduce DPC 5247 numbers in a mixed biofilm as expected since it reduced DPC 5237 in a single species biofilm. The reduction by Phage K of biofilms has been previously seen (53, 54). N1M2 alone could not reduce N1 numbers in a mixed biofilm. This was expected as N1M2 alone could not reduce N1 numbers in a single species biofilm under the same conditions. In XTT assays a combination of N1M2 and Phage K was more effective than Phage K alone but this was not seen when quantifying the biofilm. A discrepancy between XTT results and quantification was previously seen in a study of the effect of antibiotics on biofilms (55). For some strains rifampicin significantly reduced cellular viability in a biofilm as measured by plating but not by XTT and vice-versa. In a study of the antimicrobial activity of plant extracts *Bacillus subtilis* growth was reduced by 100% according to XTT assays but only 60% by plating (56). XTT values were from two separate experiments while viable count values were from three separate experiments.

In conclusion, a novel bacteriophage was isolated against a clinically important bacterial pathogen that is of great interest in the fight against antimicrobial resistance. Phage N1M2 was characterised as a jumbo phage with little homology to other *K. aerogenes* phages and jumbo phages. The most related phage was the jumbo phage *Pseudomonas* phage OBP. Interest is growing in the use of phages in place of antibiotics. *K. aerogenes* can be found as a constituent of biofilms on medical devices. Phage N1M2 was found to be effective at reducing preformed biofilms but not in stopping the formation of biofilms. Biofilms on medical devices often contain mixed bacterial communities. *S. aureus* DPC 5247 was used in biofilms in combination with *K. aerogenes* N1. Phage N1M2 in combination with Phage K were significantly better at reducing a pre-formed *K. aerogenes* and *S. aureus* biofilm than either phage alone.

References

1. **Wommack KE, Hill RT, Kessel M, Russek-Cohen E, Colwell RR.** 1992. Distribution of viruses in the Chesapeake Bay. Applied and Environmental Microbiology **58:**2965-2970.

2. **Shkoporov AN, Hill C.** 2019. Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. Cell Host Microbe **25:**195-209.

3. **Mills S, Shanahan F, Stanton C, Hill C, Coffey A, Ross RP.** 2013. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut Microbes **4:**4-16.

4. **Norman JM, Handley SA, Baldridge MT.** 2015. Disease–specific alterations in the enteric virome in inflammatory bowel disease. Cell **160**.

5. **Guang-Han O, Leang-Chung C, Vellasamy KM, Mariappan V, Li-Yen C, Vadivelu J.** 2016. Experimental Phage Therapy for *Burkholderia pseudomallei* Infection. PLoS ONE **11:**e0158213.

6. **Bai J, Kim Y-T, Ryu S, Lee J-H.** 2016. Biocontrol and Rapid Detection of Food-borne Pathogens Using Bacteriophages and Endolysins. Frontiers in Microbiology **7:**474.

7. **Turnbough Jr CL.** 2003. Discovery of phage display peptide ligands for species-specific detection of *Bacillus* spores. Journal of Microbiological Methods **53:**263-271.

8. **Pogue JM, Kaye KS, Cohen DA, Marchaim D.** 2015. Appropriate antimicrobial therapy in the era of multidrug-resistant human pathogens. Clinical Microbiology and Infection **21:**302-312.

9. **World Health Organization.** 2017. WHO Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics.

10. **Munson E, Carroll KC.** 2019. An Update on the Novel Genera and Species and Revised Taxonomic Status of Bacterial Organisms Described in 2016 and 2017. J Clin Microbiol **57**.

11. **Hede K.** 2014. Antibiotic resistance: An infectious arms race. Nature **509:**S2.

12. **Kutateladze M, Adamia R.** 2008. Phage therapy experience at the Eliava Institute. Médecine et Maladies Infectieuses **38:**426-430.

13. **Jault P, Leclerc T, Jennes S, Pirnay JP, Que Y-A, Resch G, Rousseau AF, Ravat F, Carsin H, Le Floch R, Schaal JV, Soler C, Fevre C, Arnaud I, Bretaudeau L, Gabard J.** 2018. Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by *Pseudomonas aeruginosa* (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. The Lancet Infectious Diseases **19:**35-45.

14. **Wright A, Hawkins CH, Änggård EE, Harper DR.** 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. Clinical Otolaryngology **34:**349-357.

15. **Sillankorva SM, Oliveira H, Azeredo J.** 2012. Bacteriophages and Their Role in Food Safety. International Journal of Microbiology **2012:**863945.

16. **Xu J, Chen M, He L, Zhang S, Ding T, Yao H, Lu C, Zhang W.** 2015. Isolation and characterization of a T4-like phage with a relatively wide host

range within *Escherichia coli*. J Basic Microbiol doi:10.1002/jobm.201500440.

17. **Hwang J-Y, Kim J-E, Song Y-J, Park J-H.** 2016. Safety of using *Escherichia coli* bacteriophages as a sanitizing agent based on inflammatory responses in rats. Food Science and Biotechnology **25:**355-360.

18. **Jamalludeen N, Johnson RP, Friendship R, Kropinski AM, Lingohr EJ, Gyles CL.** 2007. Isolation and characterization of nine bacteriophages that lyse O149 enterotoxigenic *Escherichia coli*. Veterinary Microbiology **124:**47-57.

19. **Kim SG, Jun JW, Giri SS, Yun S, Kim HJ, Chi C, Kim SW, Park SC.** 2017. Complete Genome Sequence of *Staphylococcus aureus* Bacteriophage pSa-3. Genome Announcements **5**.

20. **Dalmasso M, de Haas E, Neve H, Strain R, Cousin FJ, Stockdale SR, Ross RP, Hill C.** 2015. Isolation of a Novel Phage with Activity against *Streptococcus mutans* Biofilms. PLOS ONE **10:**e0138651.

21. **Chang H-C, Chen C-R, Lin J-W, Shen G-H, Chang K-M, Tseng Y-H, Weng S-F.** 2005. Isolation and Characterization of Novel Giant *Stenotrophomonas maltophilia* Phage φSMA5. Applied and Environmental Microbiology **71:**1387-1393.

22. **Gong Z, Wang M, Yang Q, Li Z, Xia J, Gao Y, Jiang Y, Meng X, Liu Z, Yang D, Zhang F, Shao H, Wang D.** 2017. Isolation and Complete Genome Sequence of a Novel *Pseudoalteromonas* Phage PH357 from the Yangtze River Estuary. Current Microbiology doi:10.1007/s00284-017-1244-8**:**1-8.

23. **Borriss M, Helmke E, Hanschke R, Schweder T.** 2003. Isolation and characterization of marine psychrophilic phage-host systems from Arctic sea ice. Extremophiles **7:**377-384.

24. **Sauder AB, Quinn MR, Brouillette A, Caruso S, Cresawn S, Erill I, Lewis L, Loesser-Casey K, Pate M, Scott C, Stockwell S, Temple L.** 2016. Genomic characterization and comparison of seven *Myoviridae* bacteriophage infecting *Bacillus thuringiensis*. Virology **489:**243-251.

25. **Brussow H, Fremont M, Bruttin A, Sidoti J, Constable A, Fryder V.** 1994. Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. Applied and Environmental Microbiology **60:**4537-4543.

26. **Cornelissen A, Hardies SC, Shaburova OV, Krylov VN, Mattheus W, Kropinski AM, Lavigne R.** 2012. Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse PhiKZ-related phages. J Virol **86:**1844-1852.

27. **Percival SL, Suleman L, Vuotto C, Donelli G.** 2015. Healthcare-associated infections, medical devices and biofilms: risk, tolerance and control. Journal of Medical Microbiology **64:**323-334.

28. **Kaur S, Harjai K, Chhibber S.** 2016. In Vivo Assessment of Phage and Linezolid Based Implant Coatings for Treatment of Methicillin Resistant *S. aureus* (MRSA) Mediated Orthopaedic Device Related Infections. PLoS One **11:**e0157626.

29. **Yilmaz C, Colak M, Yilmaz BC, Ersoz G, Kutateladze M, Gozlugol M.** 2013. Bacteriophage therapy in implant-related infections: an experimental study. J Bone Joint Surg Am **95:**117-125.

30. **Mulzer J, Trampuz A, Potapov EV.** 2019. Treatment of chronic left ventricular assist device infection with local application of bacteriophages. European Journal of Cardio-Thoracic Surgery doi:10.1093/ejcts/ezz295.

31. **Tkhilaishvili T, Winkler T, Muller M, Perka C, Trampuz A.** 2019. Bacteriophages as adjuvant to antibiotics for the treatment of periprosthetic joint infection caused by multidrug-resistant *Pseudomonas aeruginosa*. Antimicrob Agents Chemother doi:10.1128/aac.00924-19.

32. **Hola V, Ruzicka F, Horka M.** 2010. Microbial diversity in biofilm infections of the urinary tract with the use of sonication techniques. FEMS Immunol Med Microbiol **59:**525-528.

33. **Storti A, Pizzolitto AC, Pizzolitto EL.** 2005. Detection of mixed microbial biofilms on central venous catheters removed from Intensive care Unit Patients. Brazilian Journal of Microbiology **36:**275-280.

34. **O'Flaherty S, Ross RP, Meaney W, Fitzgerald GF, Elbreki MF, Coffey A.** 2005. Potential of the Polyvalent Anti-*Staphylococcus* Bacteriophage K for Control of Antibiotic-Resistant Staphylococci from Hospitals. Applied and Environmental Microbiology **71:**1836-1842.

35. **Łubowska N, Grygorcewicz B, Kosznik-Kwaśnicka K, Zauszkiewicz-Pawlak A, Węgrzyn A, Dołęgowska B, Piechowicz L.** 2019. Characterization of the Three New Kayviruses and Their Lytic Activity Against Multidrug-Resistant Staphylococcus aureus. Microorganisms **7:**471.

36. **Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ.** 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolecular Detection and Quantification **3:**1-8.

37. **Bornet C, Chollet R, Malléa M, Chevalier J, Davin-Regli A, Pagès J-M, Bollet C.** 2003. Imipenem and expression of multidrug efflux pump in *Enterobacter aerogenes*. Biochemical and Biophysical Research Communications **301:**985-990.

38. **Yuan Y, Gao M.** 2017. Jumbo Bacteriophages: An Overview. Frontiers in microbiology **8:**403-403.

39. **Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, Robben J, Mesyanzhinov V, Krylov VN, Volckaert G.** 2005. Genome Comparison of *Pseudomonas aeruginosa* Large Phages. Journal of Molecular Biology **354:**536-545.

40. **Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaert G.** 2002. The genome of bacteriophage φKZ of *Pseudomonas aeruginosa*. Journal of Molecular Biology **317:**1-19.

41. **Mishra CK, Choi TJ, Kang SC.** 2012. Isolation and characterization of a bacteriophage F20 virulent to *Enterobacter aerogenes*. J Gen Virol **93:**2310-2314.

42. **Li E, Wei X, Ma Y, Yin Z, Li H, Lin W, Wang X, Li C, Shen Z, Zhao R, Yang H, Jiang A, Yang W, Yuan J, Zhao X.** 2016. Isolation and characterization of a bacteriophage phiEap-2 infecting multidrug resistant *Enterobacter aerogenes*. Sci Rep **6:**28338.

43. **Zhao J, Zhang Z, Tian C, Chen X, Hu L, Wei X, Li H, Lin W, Jiang A, Feng R, Yuan J, Yin Z, Zhao X.** 2019. Characterizing the Biology of Lytic Bacteriophage vB_EaeM_φEap-3 Infecting Multidrug-Resistant *Enterobacter aerogenes*. Frontiers in Microbiology **10**.

44. **Verthe K, Possemiers S, Boon N, Vaneechoutte M, Verstraete W.** 2004. Stability and activity of an *Enterobacter aerogenes*-specific bacteriophage under simulated gastro-intestinal conditions. Appl Microbiol Biotechnol **65:**465-472.

45. **Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen KV, Nguyen TV, Dao TT, Mensink M, Minh VL, Nhu NT, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR.** 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci U S A **112:**E3574-3581.

46. **Chadha P, Katare OP, Chhibber S.** 2016. In vivo efficacy of single phage versus phage cocktail in resolving burn wound infection in BALB/c mice. Microbial Pathogenesis **99:**68-77.

47. **Chhibber S, Kaur S, Kumari S.** 2008. Therapeutic potential of bacteriophage in treating *Klebsiella pneumoniae* B5055-mediated lobar pneumonia in mice. Journal of Medical Microbiology **57:**1508-1513.

48. **Nir-Paz R, Gelman D, Khouri A, Sisson BM, Fackler J, Alkalay-Oren S, Khalifa L, Rimon A, Yerushalmy O, Bader R, Amit S, Coppenhagen-**

Glazer S, Henry M, Quinones J, Malagon F, Biswas B, Moses AE, Merril G, Schooley RT, Brownstein MJ, Weil YA, Hazan R. 2019. Successful treatment of antibiotic resistant poly-microbial bone infection with bacteriophages and antibiotics combination. Clinical Infectious Diseases doi:10.1093/cid/ciz222.

49. **Ceyssens P-J, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, Blasdel B, De Smet J, Noben J-P, Bläsi U, Severinov K, Lavigne R.** 2014. Development of Giant Bacteriophage ϕKZ Is Independent of the Host Transcription Apparatus. Journal of Virology **88:**10501-10510.

50. **Yoshikawa G, Askora A, Blanc-Mathieu R, Kawasaki T, Li Y, Nakano M, Ogata H, Yamada T.** 2018. *Xanthomonas citri* jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. Scientific Reports **8:**4486.

51. **Hosseinidoust Z, Tufenkji N, van de Ven TGM.** 2013. Formation of biofilms under phage predation: considerations concerning a biofilm increase. Biofouling **29:**457-468.

52. **González S, Fernández L, Campelo AB, Gutiérrez D, Martínez B, Rodríguez A, García P.** 2017. The Behavior of *Staphylococcus aureus* Dual-Species Biofilms Treated with Bacteriophage phiIPLA-RODI Depends on the Accompanying Microorganism. Applied and Environmental Microbiology **83:**e02821-02816.

53. **Lungren MP, Christensen D, Kankotia R, Falk I, Paxton BE, Kim CY.** 2013. Bacteriophage K for reduction of *Staphylococcus aureus* biofilm on central venous catheter material. Bacteriophage **3:**e26825.

54. **Cerca N, Oliveira R, Azeredo J.** 2007. Susceptibility of *Staphylococcus epidermidis* planktonic cells and biofilms to the lytic action of staphylococcus bacteriophage K. Letters in Applied Microbiology **45:**313-317.

55. **Cerca N, Martins S, Cerca F, Jefferson KK, Pier GB, Oliveira R, Azeredo J.** 2005. Comparative assessment of antibiotic susceptibility of coagulase-negative staphylococci in biofilm versus planktonic culture as assessed by bacterial enumeration or rapid XTT colorimetry. Journal of Antimicrobial Chemotherapy **56:**331-336.

56. **Al-Bakri AG, Afifi FU.** 2007. Evaluation of antimicrobial activity of selected plant extracts by rapid XTT colorimetry and bacterial enumeration. Journal of Microbiological Methods **68:**19-25.

57. **Sullivan MJ, Petty NK, Beatson SA.** 2011. Easyfig: a genome comparison visualizer. Bioinformatics (Oxford, England) **27:**1009-1010.

58. **Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS.** 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research **44:**W16-W21.

59. **Chen IMA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A, Huang J, Andersen E, Huntemann M, Varghese N, Hadjithomas M, Tennessen K, Nielsen T, Ivanova NN, Kyrpides NC.** 2017. IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic acids research **45:**D507-D516.

60. **Nurk S, Meleshko D, Korobeynikov A, Pevzner PA.** 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res **27:**824-834.

61. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

62. **González-Tortuero E, Sutton TDS, Velayudhan V, Shkoporov AN, Draper LA, Stockdale SR, Ross RP, Hill C.** 2018. VIGA: a sensitive, precise and automatic *de novo* VIral Genome Annotator. bioRxiv doi:10.1101/277509**:**277509.

63. **Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S.** 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics (Oxford, England) **30:**1236-1240.

64. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-580.

65. **Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A.** 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci **12:**1652-1662.

66. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955-964.

67. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res **32:**11-16.

68. **Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D.** 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol **8:**11-13.

69. **Zuker M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research **31:**3406-3415.

70. **Wernersson R.** 2005. FeatureExtract--extraction of sequence annotation made easy. Nucleic Acids Res **33:**W567-569.

71. **Bailey TL, Elkan C.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:**28-36.

72. **Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G.** 2010. Interactive microbial genome visualization with GView. Bioinformatics **26:**3125-3126.

73. **Kumar S, Stecher G, Tamura K.** 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol **33:**1870-1874.

74. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research **32:**1792-1797.

75. **Jones DT, Taylor WR, Thornton JM.** 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8:**275-282.

76. **Felsenstein J.** 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution **39:**783-791.

77. **Göker M, Meier-Kolthoff JP.** 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics **33:**3396-3404.

78. **Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M.** 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics **14:**60.

79. **Tunney MM, Ramage G, Field TR, Moriarty TF, Storey DG.** 2004. Rapid colorimetric assay for antimicrobial susceptibility testing of *Pseudomonas aeruginosa*. Antimicrobial agents and chemotherapy **48:**1879-1881.

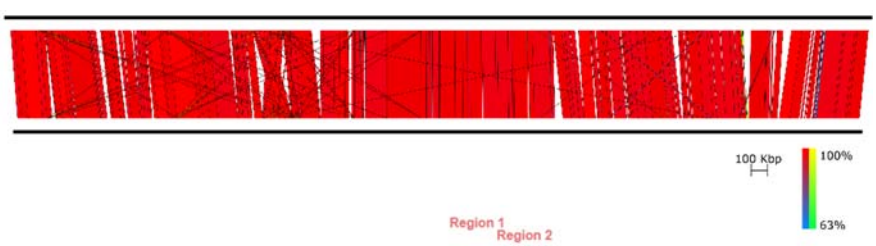
Tables and figures

Table 1. Codon usage and tRNA present in phage N1M2 and N1. Codon usage was determined using the University of Georgia online amino acid and codon usage statictics online tool. The presence of transfer RNA genes was determined using tRNAscan-SE and ARAGORN.

	Codon	N1M2 frequency %	N1 frequency %	Ratio (phage/host)	N1M2 tRNA (location)	N1 tRNA (location)
	GCC	0.84	3.29	0.26	0	3 (569533-569608) (569682-569758) (569832-569907)
	GCG	0.57	4.13	0.14	0	0
	CGA	0.29	0.82	0.35	0	0
	CGC	0.60	3.11	0.19	0	0
	CGG	0.16	1.62	0.10	0	0
	AGG	0.17	0.38	0.45	0	0
	TGC	0.24	1.13	0.21	0	0
	CAG	1.59	2.77	0.57	0	4 (295367-295451) (295480-295563) (295593-295677) (925991-926076)
	GGC	0.84	3.84	0.22	0	0
	GGG	0.77	1.32	0.58	0	0
Phage<Host	CAC	0.74	1.11	0.67	0	0
	CAT	1.04	1.23	0.85	1 (249035-249111)	0
	CTC	0.73	1.23	0.59	0	0
	CTG	2.25	4.95	0.45	0	0
	CCC	0.16	0.83	0.19	0	0
	CCG	0.55	3.10	0.18	0	1 (926231-926307)
	TCC	0.81	1.04	0.78	0	0
	TCG	0.31	1.38	0.22	0	0
	AGC	0.62	2.08	0.30	0	0
	ACC	1.75	2.51	0.70	0	0
	ACG	0.44	1.44	0.31	0	0
	TGG	1.22	1.54	0.79	0	0
	GTC	0.94	1.74	0.54	0	1 (962722-962798)
	GTG	0.84	2.31	0.36	0	1 (926098-926174)
	GAC	1.90	1.83	1.04	0	0
	GAG	1.64	1.75	0.94	0	0
Phage=Host	GGA	0.86	0.86	1.00	0	0
	CTA	0.57	0.60	0.95	0	0
	TTT	1.88	2.08	0.90	0	0
Phage>Host	GCA	1.77	1.25	1.42	0	0
	GCT	2.92	1.63	1.79	0	0

	CGT	2.56	1.66	1.54	0	0
	AGA	0.72	0.39	1.85	0	0
	AAC	3.15	2.08	1.51	0	0
	AAT	2.68	1.41	1.90	0	0
	GAT	4.63	2.79	1.66	0	2 (786427-786503) (966444-966520)
	TGT	0.62	0.53	1.17	0	1 (778986-77906)
	CAA	1.79	1.15	1.56	0	2 (331627-331705) (440212-440130)
	GAA	4.86	2.86	1.70	0	1 (610794-610869)
	GGT	4.15	1.68	2.47	0	1 (778693-778620)
	ATA	0.71	0.58	1.22	0	0
	ATC	3.32	2.60	1.28	0	0
	ATT	3.08	2.15	1.43	0	0
	CTT	1.55	1.01	1.53	0	0
	TTA	1.62	1.02	1.59	0	0
	TTG	1.45	1.27	1.14	0	1 (925945-925873)
	AAA	4.40	2.82	1.56	0	0
	AAG	2.44	1.30	1.88	0	0
	ATG	2.49	2.16	1.15	0	0
	TTC	2.53	1.84	1.38	0	4 (409621-409696) (740942-741019) (778699-778773) (875919-875994)
	CCA	1.58	0.73	2.16	0	1 (962668-962592)
	CCT	1.83	0.70	2.61	0	0
	TCA	0.88	0.78	1.13	0	0
	TCT	2.13	0.73	2.92	1 (249512-249588)	0
	AGT	1.23	0.58	2.12	0	0
	ACA	1.25	0.48	2.6	0	0
	ACT	3.02	0.68	4.44	0	0
	TAC	1.43	1.23	1.16	0	0
	TAT	2.69	1.50	1.79	0	0
	GTA	1.77	0.94	1.88	0	1 (778891-778975)
	GTT	3.40	1.43	2.38	1 (249909-249982)	0

A.



B.

C.

Klebsiella aerogenes KCTC 2190 encoded functions:

Pfam categories	Count	Percent
Amino acid transport and metabolism	201	8.46
Carbohydrate transport and metabolism	203	8.54
Cell cycle control, cell division, chromosome partitioning	24	1.01
Cell motility	84	3.53
Cell wall/membrane/envelope biogenesis	169	7.11
Chromatin structure and dynamics	1	0.4
Coenzyme transport and metabolism	119	5.01
Defense mechanisms	11	0.46
Energy production and conversion	181	7.61
Extracellular structures	1	0.04
Function unknown	279	11.74
General function prediction only	186	7.82
Inorganic ion transport and metabolism	143	6.02
Intracellular trafficking, secretion, and vesicular transport	102	4.29
Lipid transport and metabolism	48	2.02
Nucleotide transport and metabolism	54	2.27
Posttranslational modification, protein turnover, chaperones	77	3.24
Replication, recombination and repair	104	4.38
Secondary metabolites biosynthesis, transport and catabolism	30	1.26
Signal transduction mechanisms	59	2.48
Transcritption	119	5.01
Translation, ribosomal structure and biogenesis	182	7.66

Defense mechanisms include:

Locus tag	Pfam name	Function
EAE_05815	beta-lactamase	pfam00144
EAE_12120	beta-lactamase/D-alanine carboxypeptidase	pfam00144
EAE_12325	beta-lactam binding protein AmpH	pfam00144
EAE_17595	putative chloramphenicol acetyltranferase	pfam00302

Figure 1. A. Comparison of N1 draft genome and *K. aerogenes* KCTC 2190 B. Prophages predicted using PHASTER in N1 and *K. aerogenes* KTCT 2190 C. Pfam predicted functions of *K. aerogenes* KTCT 2190 genes.

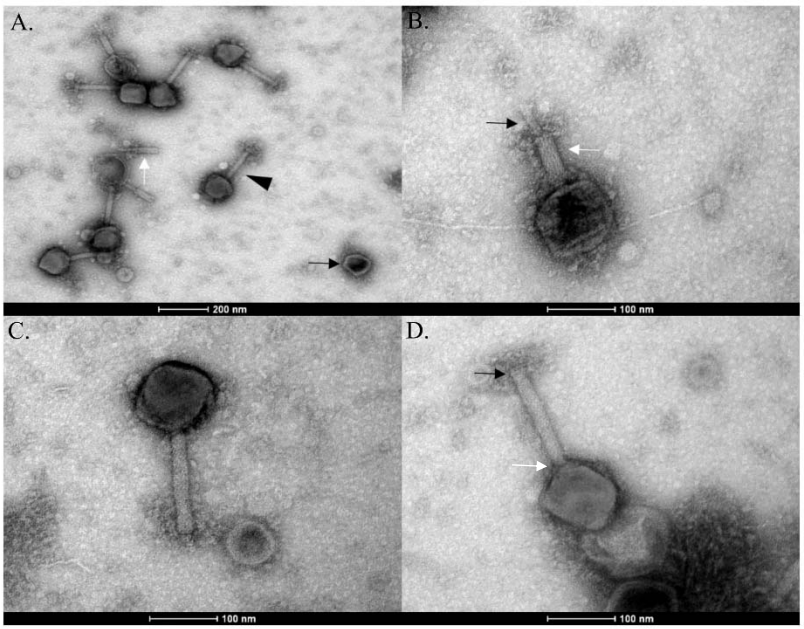


Figure 2. EMs of phage N1M2. A. Empty phage capsid (black arrow), broken phage tail (white arrow), and fully formed phage particle (black triangle). B. Contracted tail sheath (white arrow) with tail core (black arrow). C. Phage N1M2 with uncontracted tail. D. Baseplate (black arrow) and neck structure (white arrow) of uncontracted tail.

Table 2. Predicted promoter of phage N1M2 compared to predicted promoters of Pseudomonas phage OBP, Pseudomonas phage 201phi2-1, Pseudomonas phage EL and Pseudomonas virus phiKZ. Bases coloured red match the sequence of the predicted promoter of OBP promoter. Bases coloured blue match the sequence of the predicted promoter of phiKZ.

Phage	Promoter sequence
N1M2	TBYAWWWWWTTTCARRYAKATATTATYWAAGTGWA
OBP	BSHAWWWWWTTTYARRYAKATATTATYWWADTG
2012-1	TTAWTAVAA_HYWTTTRARR_BTATATTACDWHDGTG
EL	WTTTYAAACCTACATTATY
phiKZ	TATATTAC

Table 3. Host range of phage N1M2. Efficiency of plaquing is represented as a fraction with SEM of 3 separate experiments. If no N1M2 plaques formed on a strain efficiency of plaquing is represented by (-).

Strain	Efficiency of plaquing
Klebsiella aerogenes N1	1 ± 0
Klebsiella aerogenes NCIMB 10102	0.667 ± 0.17
Klebsiella pneumoniae NCIMB 13218	0.017 ± 0.009
Enterobacter gergoviae DPC 6436	-
Enterobacter cloacae DPC 6437	-
Pseudomonas aeruginosa PAO1	-
Escherichia coli 0157 ΔStx	-
Escherichia coli APC 106	-
Acinetobacter johnsonii	-
Staphylococcus aureus DPC 5247	-

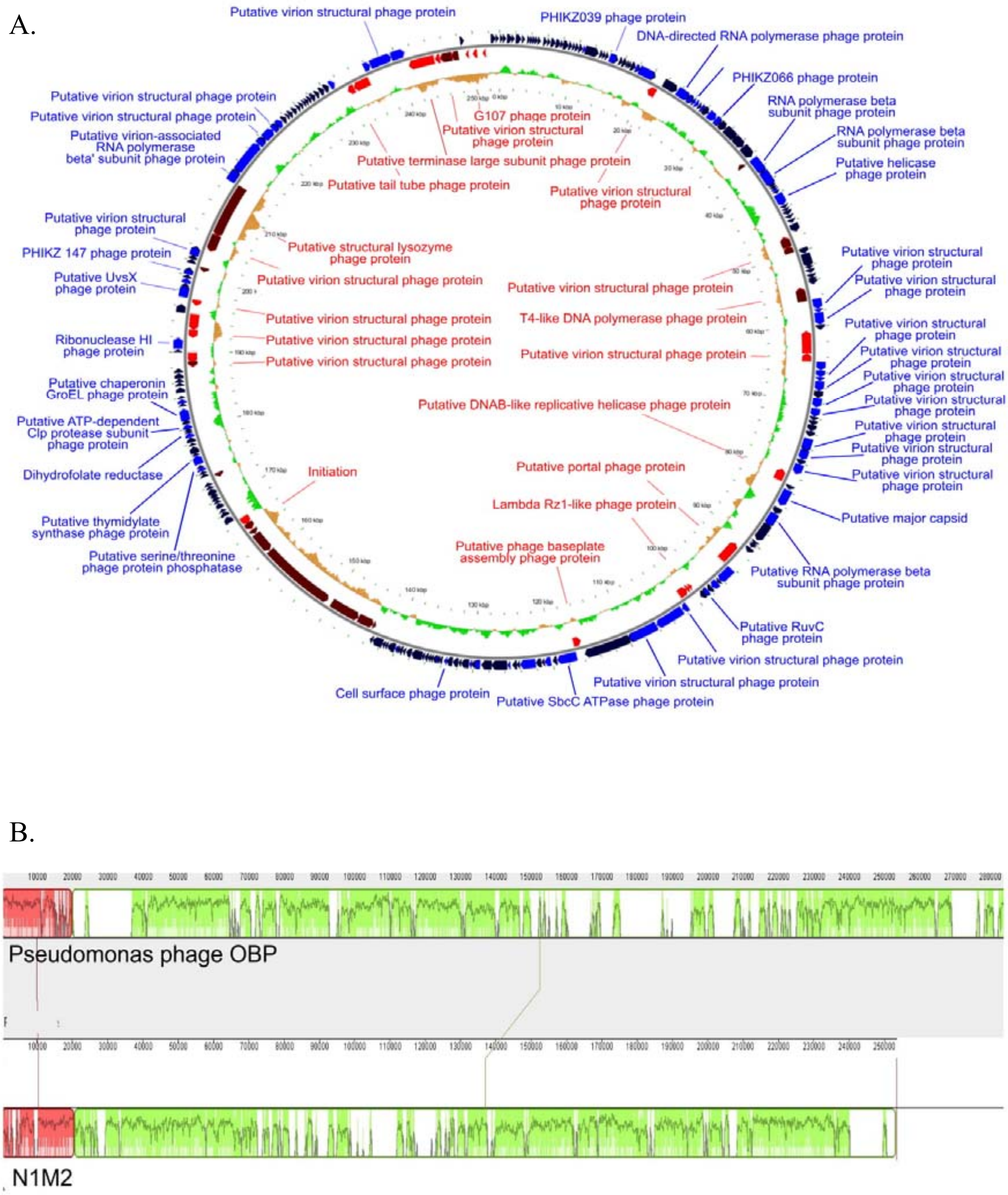


Figure 3. A. Genome map of phage N1M2. On the outer ring blue arrows represent genes on the forward strand, red arrows represent genes on the reverse strand, and black arrows represent hypothetical proteins. The inner ring represents GC skew compared to the average GC content of phage N1M2. B. Mauve alignment of phage N1M2 and Pseudomonas phage OBP.

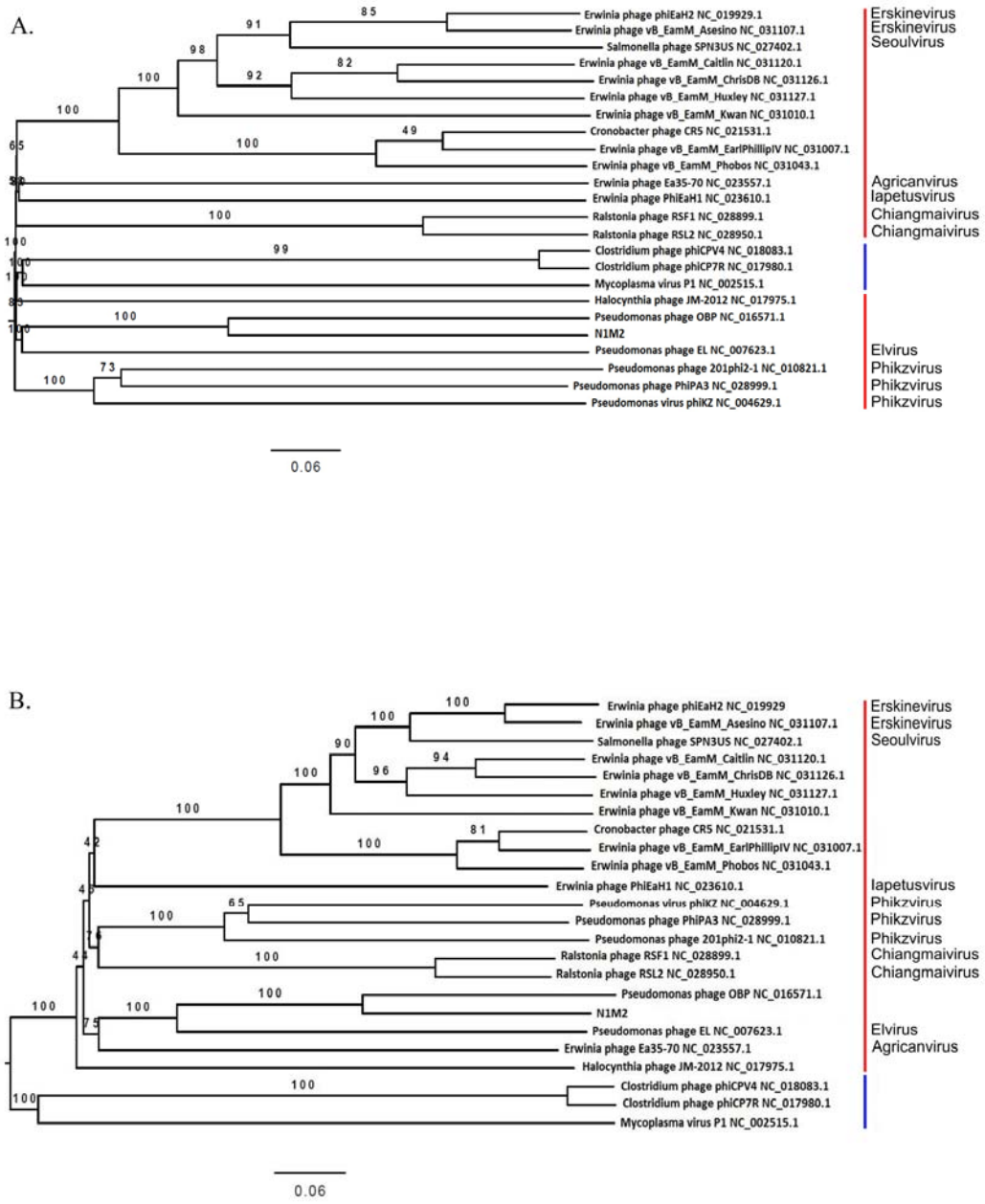


Figure 4. Victor-generated phylogenomic Genome-BLAST Distance Phylogeny (GBDP) trees of *K. aerogenes* phage N1M2 and other jumbo phages inferred A. using nucleotide identity and the formula D6 and yielding average support of 88%. B. using amino acid identity and the formula D6 and yielding average support of 86%. The numbers above branches are GBDP pseudobootstrap support values from 100 replicates. Red bars represent family *Myoviridae*, blue bars represent family *Podoviridae*.

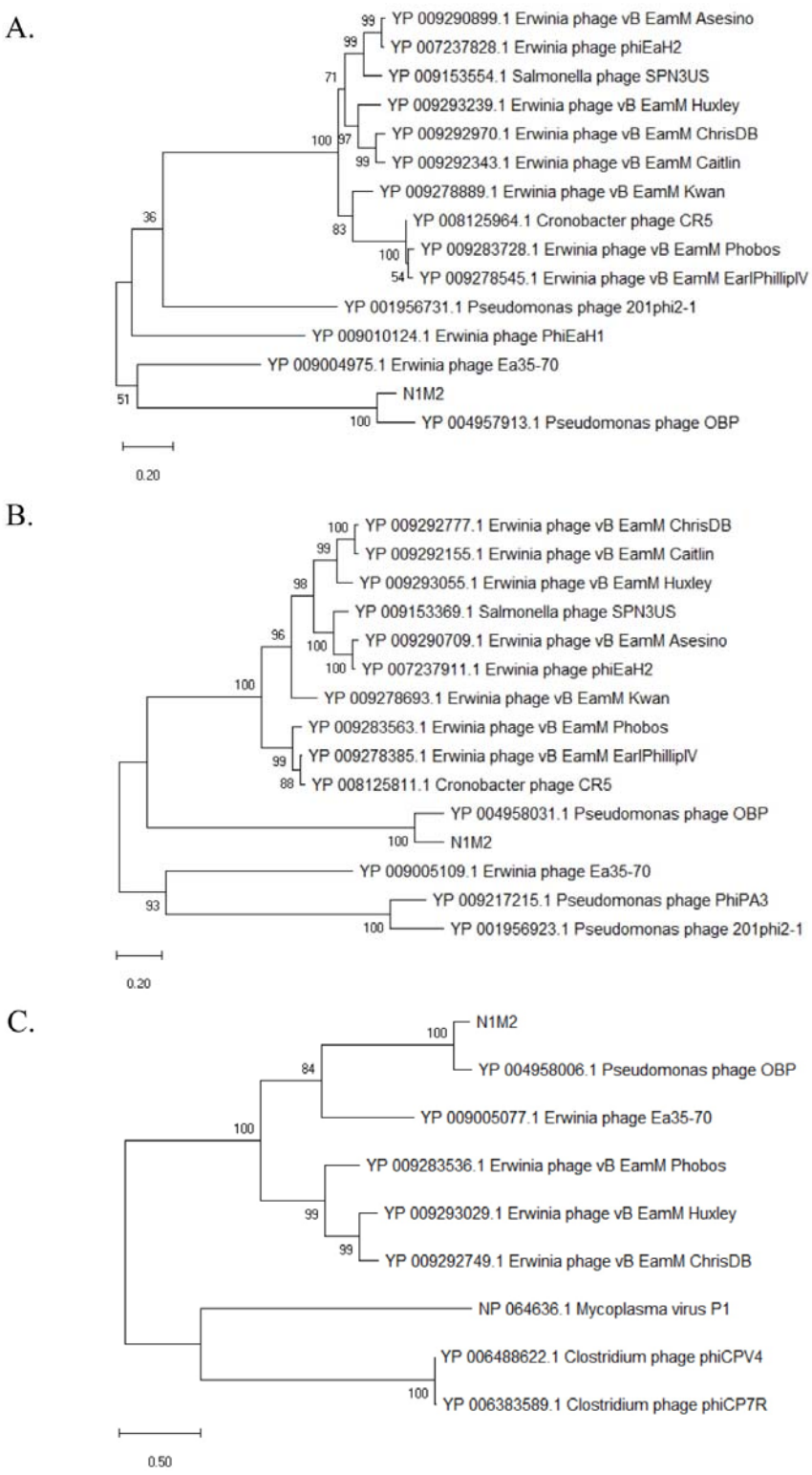


Figure 5. Phylogenetic analysis of amino acid sequences of *K. aerogenes* phage N1M2 and other jumbo phages using maximum likelihood (Whelan and Goldman substitution model), with 1000 bootstrap replicates of A. major capsid protein B. large terminase and C. DNA polymerase.

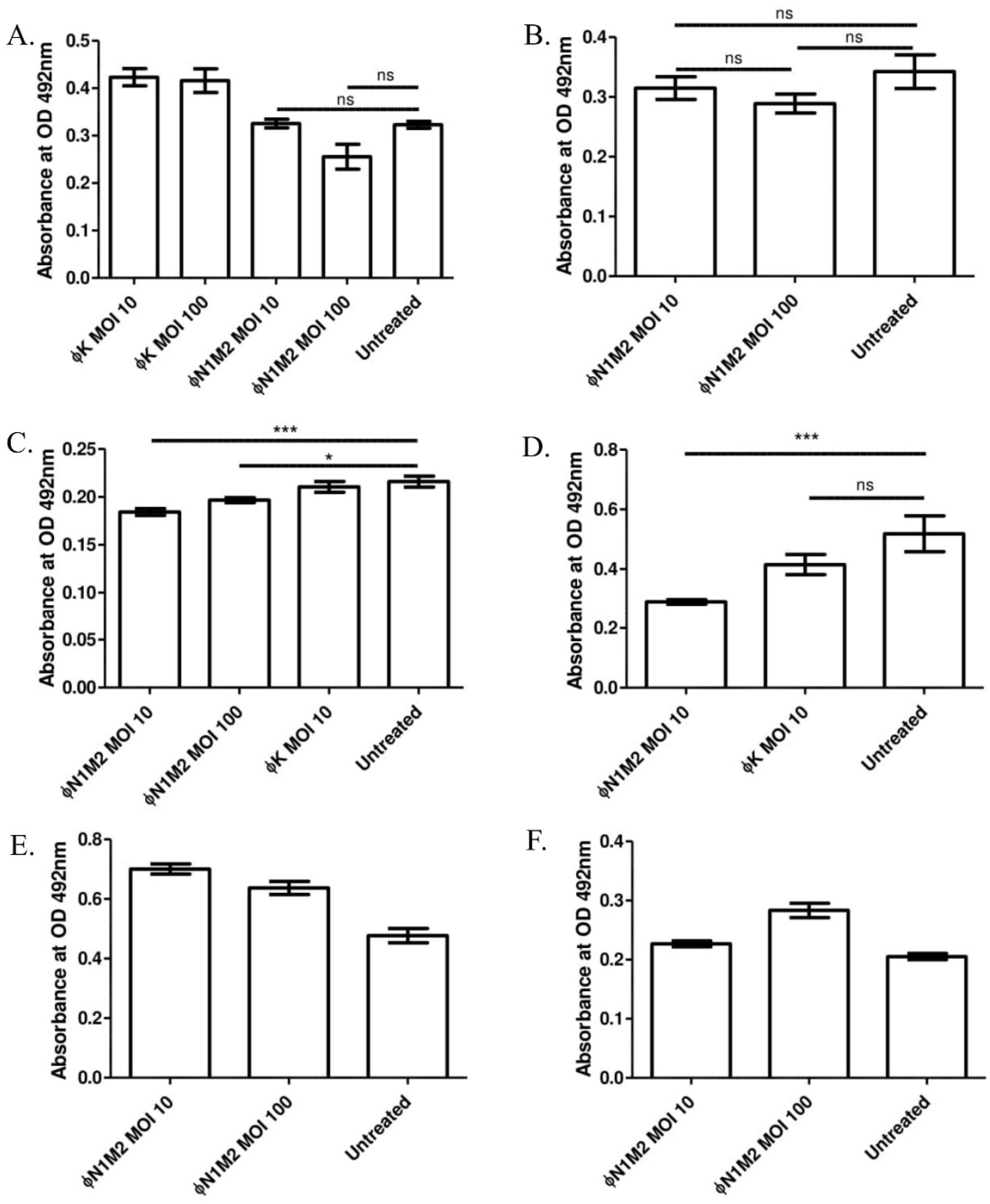


Figure 6. Effect of phage N1M2 on biofilms formed by N1 *K. aerogenes.* Biofilm metabolic activity was assessed by OD492nm measures after treatment with XTT supplemented with menadione. A. 1% glucose 48 hr biofilm, 48 hr phage B. No glucose 48 hr biofilm, 48 hr phage C. No glucose 48 hr biofilm, 72 hr phage D. No glucose 24 hr biofilm, 48 hr phage E. 1% glucose stop biofilm formation over 48 hr F. No glucose stop biofilm formation over 48 hr.

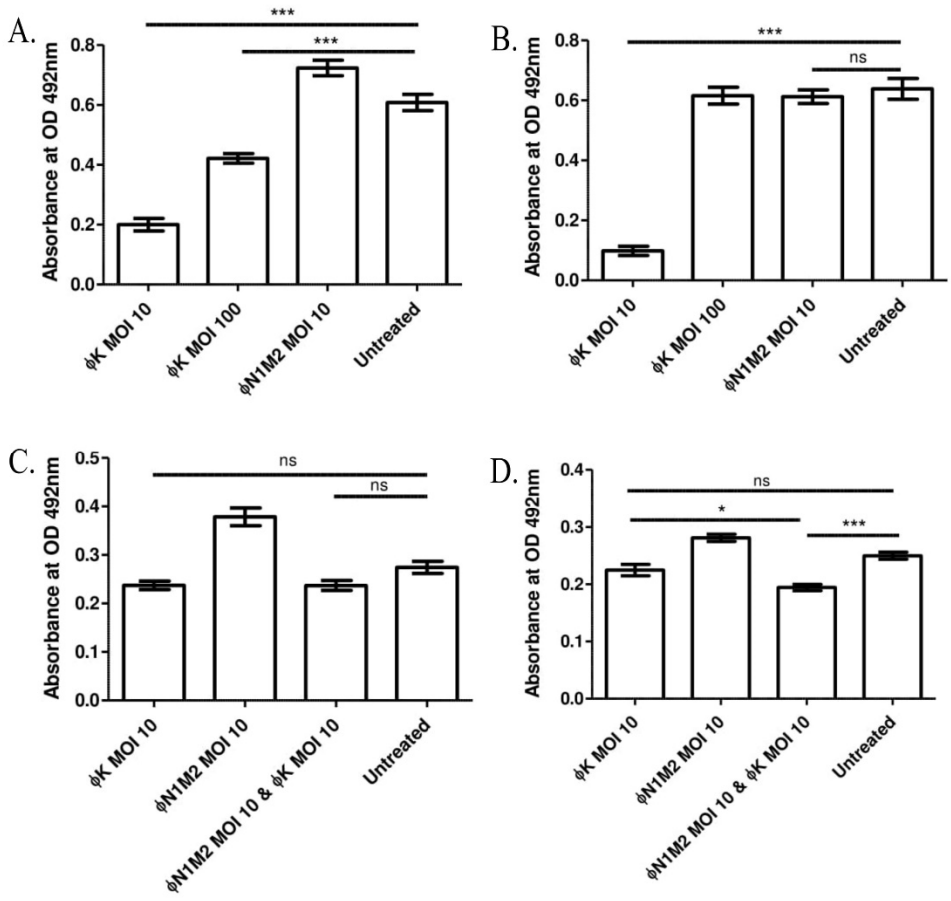


Figure 7. Biofilm metabolic activity was assessed by OD492nm measures after treatment with XTT supplemented with menadione. A. Effect of Phage K on biofilms formed by DPC 5247 *S. aureus*. 1% glucose 48 hr biofilm, 48 hr phage B. Effect of Phage K on biofilms formed by DPC 5247 *S. aureus*. No glucose 48 hr biofilm, 48 hr phage C. Effect of phage N1M2 and Phage K on mixed biofilms formed by N1 *K. aerogenes* and DPC 5247 *S.aureus*. 1% glucose 48 hr biofilm, 48 hr phage D. Effect of phage N1M2 and Phage K on mixed biofilms formed by N1 *K. aerogenes* and DPC 5247 *S.aureus.* No glucose 48 hr biofilm, 48 hr phage.

Supplementary table 1. A bank of known strains was selected to be used for phage screening.

Strain designation	Other designations	Strain
APC 53	DPC 6085, NCIMB700577, NCDO0577	*Bacillus cereus*
APC 54	DPC 6086, NCIMB700578, NCDO0578	*Bacillus cereus*
APC 55	DPC 6087, NCIMB700579, NCDO0579	*Bacillus cereus*
APC 56	DPC 6088, NCIMB700827, NCDO0827	*Bacillus cereus*
APC 57	DPC 6089, NCIMB8079, ATCC7004	*Bacillus cereus*
APC 58	DPC 6334, Bel 17Bc	*Bacillus cereus*
APC 59	DPC 6336, Bel Bc33	*Bacillus cereus*
DPC 8079		*Bacillus cereus*
APC 1762	UCC 5002	*Bacillus subtilis*
APC 1640		*Enterococcus casseliflavus*
APC 1039	EC618	*Enterococcus faecalis*
APC 1749	DPC 5152	*Enterococcus faecalis*
APC 1025	EC251	*Enterococcus faecium*
APC 1026	EC289	*Enterococcus faecium*
APC 1029	EC300	*Enterococcus faecium*
APC 1030	EC357	*Enterococcus faecium*
APC 1031	EC520	*Enterococcus faecium*
APC 1032	EC533	*Enterococcus faecium*
APC 1035	EC548	*Enterococcus faecium*
APC 1036	EC562	*Enterococcus faecium*
APC 1038	EC587	*Enterococcus faecium*
APC 1043	*EC725*	*Enterococcus faecium*
APC 1044	*EC748*	*Enterococcus faecium*
APC 1641		*Enterococcus faecium*
Lactobacillus salivarius 6482		*Lactobacillus salivarius*
APC 317	DPC 5245	*S. aureus*
APC 993	ST290	*S. aureus*
APC 994	ST291	*S. aureus*
APC 995	ST295	*S. aureus*
APC 996	ST299	*S. aureus*
APC 998	ST355	*S. aureus*
APC 1003	ST528	*S. aureus*
APC 1004	ST530	*S. aureus*
APC 1007	ST535	*S. aureus*
APC 1009	ST544	*S. aureus*
APC 1010	ST550	*S. aureus*
APC 1012	35197	*S. aureus*
APC 1019	25949	*S. aureus*
APC 1022	Newman	*S. aureus*
APC 1024	RF122	*S. aureus*
APC 1055	LMG14694, ATCC 13813, CCRC 10787, CCUG 4208, CIP 103227, DSM 2134, JCM 5671, NCFB 1348, NCTC 8181	*Streptococcus agalactiae*
APC 1759		*Streptococcus agalactiae*
APC 1755		*Streptococcus dysgalactiae*
APC 119	DPC 6143, 4001	*Streptococcus mutans*
APC 120	DC 6144, 4021	*Streptococcus mutans*
APC 121	DPC 6145, 4070	*Streptococcus mutans*
APC 122	DPC 6150, 4009	*Streptococcus mutans*
APC 123	DPC 6151, 4030	*Streptococcus mutans*

APC 124	DPC 6152, 4037	*Streptococcus mutans*
APC 125	DPC 6153, 4039	*Streptococcus mutans*
APC 126	DPC 6154, 4040	*Streptococcus mutans*
APC 127	DPC 6155, 4055	*Streptococcus mutans*
APC 128	DPC 6156, 4058	*Streptococcus mutans*
APC 129	DPC 6157, 3007	*Streptococcus mutans*
APC 130	DPC 6158, 3013	*Streptococcus mutans*
APC 131	DPC 6159, 3017	*Streptococcus mutans*
APC 132	DPC 6160, 1038	*Streptococcus mutans*
APC 133	DPC 6161, 1054	*Streptococcus mutans*
APC 134	DPC 6162, NCTC10449, ATCC 25175, SIMS, DSM 20523, ATCC 25175, IFO 13955, NCDO 2062	*Streptococcus mutans*
APC 135	DPC 6543	*Streptococcus mutans*
APC 1756	DSM 2071	*Streptococcus pyogenes*
APC 1757	DSM11728	*Streptococcus pyogenes*
APC 1758	NCDO2381	*Streptococcus pyogenes*
APC 104	DPC 6009	*Escherichia coli*
APC 105	DPC 6050	*Escherichia coli*
APC 106	DPC 6051	*Escherichia coli*
APC 109	DPC 6054, P1432	*Escherichia coli*
APC 110	DPC 6055, AR12900	*Escherichia coli*
APC 115	DPC 6472	*Escherichia coli*
APC 1220	HM605	*Escherichia coli*
Escherichia coli 042		*Escherichia coli*
Escherichia coli Nissle		*Escherichia coli*
Esherichia coli UTI89		*Escherichia coli*
APC 1977	MG1655	*Escherichia coli*
APC 176	DPC 6452	*Salmonella typhimurium*
APC 181	DPC 6547	*Salmonella typhimurium*
APC 185	DPC 6436	*Salmonella typhimurium*

Supplementary table 2. Resistance genes present in *K. aerogenes* N1 draft genome.

Start	End	Strand	Size (bp)	Protein
366872	367006	+	135	Multidrug resistance protein MdtM
367003	367434	+	432	Multidrug resistance protein MdtM
367400	367756	+	357	Multidrug resistance protein MdtM
367741	367899	+	159	Multidrug resistance protein MdtM
375134	375493	-	360	Fosfomycin resistance protein FosA
395992	396282	+	291	Small multidrug resistance family (SMR)
396382	396648	+	267	Small multidrug resistance family (SMR)
594242	594364	+	123	RND efflux system, membrane fusion protein
594373	594609	+	237	RND efflux system, membrane fusion protein
594740	594925	+	186	RND efflux system, membrane fusion protein
595140	595373	+	234	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
595333	595560	+	228	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
595535	595828	+	294	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
595825	596058	+	234	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
596018	596341	+	324	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
596494	597003	+	510	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
596996	597163	+	168	Bacteriocin/lantibiotic efflux ABC transporter, permease/ATP-binding protein
1013381	1013539	-	159	Multidrug resistance protein MdtL
1013536	1013667	-	132	Multidrug resistance protein MdtL
1013682	1014032	-	351	Multidrug resistance protein MdtL
1014085	1014483	-	399	Multidrug resistance protein MdtL
1062821	1063504	-	684	Multidrug resistance protein EmrD
1063461	1063679	-	219	Multidrug resistance protein EmrD
1063631	1063930	-	300	Multidrug resistance protein EmrD
1092936	1093298	-	363	Drug resistance transporter EmrB/QacA subfamily
1093298	1093591	-	294	Drug resistance transporter EmrB/QacA subfamily
1096061	1096552	-	492	RND efflux system, membrane fusion protein
1096673	1096924	+	252	RND efflux system, membrane fusion protein
1096899	1097234	+	336	RND efflux system, membrane fusion protein
1097188	1097952	+	765	RND efflux system, inner membrane transporter
1097989	1098135	+	147	RND efflux system, inner membrane transporter
1098149	1098739	+	591	RND efflux system, inner membrane transporter
1098711	1099190	+	480	RND efflux system, inner membrane transporter
1099212	1099739	+	528	RND efflux system, inner membrane transporter
1099718	1100335	+	618	RND efflux system, inner membrane transporter
1232604	1232894	+	291	Putative resistance protein
1232860	1233060	+	201	Putative resistance protein

1233035	1233139	+	105	Putative resistance protein
1233139	1233561	+	423	Putative resistance protein
1233524	1233646	+	123	Putative resistance protein
1233697	1233840	+	144	Putative resistance protein
1487046	1487204	-	159	RND efflux system, inner membrane transporter CmeB
1487295	1488350	-	1056	Multidrug efflux system AcrEF-TolC, inner-membrane proton/drug antiporter AcrF (RND type)
1488379	1488789	-	411	Multidrug efflux system AcrEF-TolC, inner-membrane proton/drug antiporter AcrF (RND type)
1488756	1489157	-	402	Multidrug efflux system AcrEF-TolC, inner-membrane proton/drug antiporter AcrF (RND type)
1489157	1489447	-	291	Multidrug efflux system AcrEF-TolC, inner-membrane proton/drug antiporter AcrF (RND type)
1489545	1489910	-	366	Multidrug efflux system AcrEF-TolC, inner-membrane proton/drug antiporter AcrF (RND type)
1490001	1490222	-	222	Multidrug efflux system AcrEF-TolC, inner-membrane proton/drug antiporter AcrF (RND type)
1490236	1490601	-	366	Multidrug efflux system AcrEF-TolC, membrane fusion component AcrE
1490598	1490858	-	261	Multidrug efflux system AcrEF-TolC, membrane fusion component AcrE
1490827	1491378	-	552	Multidrug efflux system AcrEF-TolC, membrane fusion component AcrE
1943717	1944244	-	528	Multidrug resistance outer membrane protein MdtP
1944216	1944545	-	330	Multidrug resistance outer membrane protein MdtP
1944565	1944981	-	417	Multidrug resistance outer membrane protein MdtP
1977822	1978478	+	657	Permease of the drug/metabolite transporter (DMT) superfamily
2167705	2168190	-	486	Multidrug efflux system EmrAB-OMF, inner-membrane proton/drug antiporter EmrB (MFS type)
2168181	2168408	-	228	Multidrug efflux system EmrAB-OMF, inner-membrane proton/drug antiporter EmrB (MFS type)
2168384	2168506	-	123	Multidrug efflux system EmrAB-OMF, inner-membrane proton/drug antiporter EmrB (MFS type)
2168559	2168675	-	117	Multidrug efflux system EmrAB-OMF, inner-membrane proton/drug antiporter EmrB (MFS type)
2168659	2169288	-	630	Multidrug efflux system EmrAB-OMF, inner-membrane proton/drug antiporter EmrB (MFS type)
2169588	2170187	-	600	Multidrug efflux system EmrAB-OMF, membrane fusion component EmrA
2170301	2170513	-	213	Multidrug efflux system EmrAB-OMF, membrane fusion component EmrA
2170656	2170772	-	117	Multidrug resistance regulator EmrR
2170931	2171053	-	123	Multidrug resistance regulator EmrR
2171077	2171268	-	192	Multidrug resistance regulator EmrR
2206785	2207132	+	348	Multidrug resistance protein ErmB
2207174	2207566	+	393	Multidrug resistance protein ErmB
2207603	2207902	+	300	Multidrug resistance protein ErmB
2395428	2395754	-	327	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)
2395871	2396503	-	633	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)

2396461	2397096	-	636	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)
2397099	2397434	-	336	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)
2397455	2397658	-	204	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)
2397721	2398134	-	414	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)
2398107	2398478	-	372	Aminoglycosides efflux system AcrAD-TolC, inner-membrane proton/drug antiporter AcrD (RND type)
2630196	2630564	+	369	Multidrug resistance transporter - Bicyclomycin resistance protein Bcr
2630564	2630755	+	192	Multidrug resistance transporter - Bicyclomycin resistance protein Bcr
2630736	2630849	+	114	Multidrug resistance transporter - Bicyclomycin resistance protein Bcr
2630876	2631007	+	132	Multidrug resistance transporter - Bicyclomycin resistance protein Bcr
2630979	2631503	+	525	Multidrug resistance transporter - Bicyclomycin resistance protein Bcr
2685749	2686591	+	843	Putative multidrug resistance outer membrane protein MdtQ
2686554	2686691	+	138	Putative multidrug resistance outer membrane protein MdtQ
2686721	2686951	+	231	Putative multidrug resistance outer membrane protein MdtQ
2739525	2740010	-	486	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2739991	2740539	-	549	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2740586	2740711	-	126	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2740689	2740826	-	141	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2740831	2741127	-	297	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2741124	2741768	-	645	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2741807	2741935	-	129	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2741919	2742203	-	285	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2742223	2742621	-	399	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2742714	2742935	-	222	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2742932	2743837	-	906	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2743851	2744522	-	672	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2744562	2745701	-	1140	Multidrug efflux system MdtABC-TolC, inner-membrane proton/drug antiporter MdtC (RND type)
2745674	2746384	-	711	Multidrug efflux system MdtABC-TolC, membrane fusion component MdtA
2746392	2746817	-	426	Multidrug efflux system MdtABC-TolC, membrane fusion component MdtA
3408656	3408946	-	291	Permease of the drug/metabolite transporter (DMT) superfamily

3408960	3409220	-	261	Permease of the drug/metabolite transporter (DMT) superfamily
3409183	3409347	-	165	Permease of the drug/metabolite transporter (DMT) superfamily
3621902	3622261	+	360	Permease of the drug/metabolite transporter (DMT) superfamily
3622271	3622723	+	453	Permease of the drug/metabolite transporter (DMT) superfamily
3653646	3653867	+	222	Multiple antibiotic resistance protein MarR
3654038	3654274	+	237	Multiple antibiotic resistance protein MarA
3654229	3654357	+	129	Multiple antibiotic resistance protein MarA
3654414	3654563	+	150	Multiple antibiotic resistance protein MarB
3654560	3655003	-	444	Permease of the drug/metabolite transporter (DMT) superfamily
3654981	3655451	-	471	Permease of the drug/metabolite transporter (DMT) superfamily
3812304	3812795	+	492	Multidrug efflux transporter MdtK/NorM (MATE family)
3812863	3813018	+	156	Multidrug efflux transporter MdtK/NorM (MATE family)
3813088	3813321	+	234	Multidrug efflux transporter MdtK/NorM (MATE family)
3813334	3813690	+	357	Multidrug efflux transporter MdtK/NorM (MATE family)
3910878	3911066	+	189	RND efflux system, inner membrane transporter
3996022	3996360	+	339	Small multidrug resistance family
4022413	4022739	+	327	Permease of the drug/metabolite transporter (DMT) superfamily
4022724	4023047	+	324	Permease of the drug/metabolite transporter (DMT) superfamily
4023162	4023290	+	129	Permease of the drug/metabolite transporter (DMT) superfamily
4096282	4096521	-	240	RND efflux system, inner membrane transporter
4069566	4070015	-	450	RND efflux system, inner membrane transporter
4069984	4070094	-	111	RND efflux system, inner membrane transporter
4070094	4070222	-	129	RND efflux system, inner membrane transporter
4070203	4071120	-	918	RND efflux system, inner membrane transporter
4071172	4071354	-	183	RND efflux system, inner membrane transporter
4071315	4071602	-	288	RND efflux system, inner membrane transporter
4071605	4071736	-	132	RND efflux system, inner membrane transporter
4071733	4071879	-	147	RND efflux system, inner membrane transporter
4071940	4072116	-	177	RND efflux system, inner membrane transporter
4072182	4072328	-	147	RND efflux system, inner membrane transporter
4072312	4072449	-	138	RND efflux system, inner membrane transporter
4072517	4072813	-	297	RND efflux system, membrane fusion protein
4072837	4072959	-	123	RND efflux system, membrane fusion protein
4072996	4073124	-	129	RND efflux system, membrane fusion protein
4073121	4073657	-	537	RND efflux system, membrane fusion protein
4150746	4150934	+	189	Multidrug resistance protein MdtH
4150895	4151170	+	276	Multidrug resistance protein MdtH
4151131	4151865	+	735	Multidrug resistance protein MdtH
4158614	4158748	+	135	Multidrug resistance protein MdtG

4158762	4159766	+	1005	Multidrug resistance protein MdtG
4195855	4196394	-	540	Permease of the drug/metabolite transporter (DMT) superfamily
4195662	4196856	-	195	Permease of the drug/metabolite transporter (DMT) superfamily
4471180	4471380	-	201	Macrolide-specific efflux protein MacA
4471377	4472069	-	693	Macrolide-specific efflux protein MacA
4472123	4472251	-	129	Macrolide-specific efflux protein MacA
4519566	4519937	-	372	Multidrug translocase MdfA
4519948	4520121	-	174	Multidrug translocase MdfA
4520106	4520213	-	108	Multidrug translocase MdfA
4520210	4520509	-	300	Multidrug translocase MdfA
4520542	4520946	-	405	Multidrug translocase MdfA
4912248	4912634	+	387	Fosmidomycin resistance protein
4912741	4913205	+	465	Fosmidomycin resistance protein
4929836	4930288	+	453	Multidrug efflux system AcrAB-TolC, membrane fusion component AcrA
4930320	4930838	+	519	Multidrug efflux system AcrAB-TolC, membrane fusion component AcrA
4930838	4930999	+	162	Multidrug efflux system AcrAB-TolC, membrane fusion component AcrA
4931198	4931398	+	201	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4931395	4931748	+	354	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4931705	4932019	+	315	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4932016	4932243	+	228	RND efflux system, inner membrane transporter CmeB
4932251	4932913	+	663	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4932974	4933099	+	126	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4933134	4933253	+	120	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4933324	4933905	+	582	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)
4933932	4934138	+	207	Multidrug efflux system AcrAB-TolC, inner-membrane proton/drug antiporter AcrB (RND type)

Supplementary table 3. ORFS table N1M2. Genes are colour coded. Structural proteins are shown in yellow, DNA replication genes are shown in blue, DNA recombination and repair genes are shown in green, Nucleotide metabolism genes are shown in peach, and Lysis genes are shown in gold.

ORF	Start	Stop	Strand	Start codon	Stop codon	Size (aa)	pI	kDa	Function	Best match (extent, % aa identity)	Accession number	e-value	TMHMM	LipoP	InterProScan
1	51	425	+	ATG	TAA	124	5.05	14.13	Hypothetical protein				0	–	–
2	403	819	+	ATG	TAA	138	6.95	15.73	unnamed protein product	Pseudomonas phage OBP (36/139, 26%)	YP_004957926.1	2.00E-08	0	–	–
3	929	1561	+	ATG	TAG	210	7.57	24.9	Hypothetical protein				1	–	–
4	1570	1821	+	ATG	TAA	83	8.39	9.82	Hypothetical protein				1	–	–
5	2029	2724	+	ATG	TAA	231	9.21	26.12	unnamed protein product	Pseudomonas phage OBP (55/230, 24%)	YP_004957929.1	8.00E-15	1	–	–
6	2803	3195	+	ATG	TAA	130	9.13	14.38	unnamed protein product	Pseudomonas phage OBP (73/121, 60%)	YP_004957931.1	1.00E-42	1	–	–
7	3535	3798	+	ATG	TAA	87	9.43	10.05	Hypothetical protein				0	–	–
8	3809	4168	+	ATG	TAA	119	4.83	13.62	Hypothetical protein				0	–	–
9	4178	4432	+	ATG	TAA	84	4.68	9.81	Hypothetical protein				0	–	–
10	4432	4812	+	ATG	TAA	126	9.21	14.48	Hypothetical protein				1	–	–
11	4819	5175	+	ATG	TAG	118	4.55	13.74	Hypothetical protein				0	–	–
12	5503	5991	+	ATG	TAA	162	9.86	19.24	unnamed protein product	Pseudomonas phage OBP (42/135,31%)	YP_004957941.1	2.00E-07	0	–	–
13	6071	6340	+	ATG	TAA	89	7.78	10.6	Hypothetical protein				0	–	–
14	6447	7058	+	ATG	TAA	203	5.37	22.91	Hypothetical protein				0	–	–
15	7074	7706	+	ATG	TAA	210	8.82	23.93	Hypothetical protein				0	–	–
16	7706	8026	+	ATG	TGA	106	9.19	12.24	unnamed protein product	Pseudomonas phage OBP (58/106, 55%)	YP_004957943.1	3.00E-31	0	–	–
17	8023	8727	+	ATG	TGA	234	5.31	27.17	Hypothetical protein				0	–	–
18	8729	9190	+	ATG	TAA	153	5.21	17.18	Hypothetical protein				0	–	–
19	9242	9568	+	ATG	TAA	108	9.56	12.07	Hypothetical protein				0	–	–
20	9578	10033	+	ATG	TAA	151	9.37	17.01	unnamed protein product	Pseudomonas phage OBP (72/151, 48%)	YP_004957947.1	2.00E-40	0	–	–
21	10035	10538	+	ATG	TGA	167	5.54	19.1	unnamed protein product	Pseudomonas phage OBP (49/158, 31%)	YP_004957948.1	5.00E-13	0	–	–
22	10646	10879	+	ATG	TAA	77	4.61	9.05	unnamed protein product	Pseudomonas phage OBP (35/73, 48%)	YP_004957949.1	2.00E-14	0	–	–
23	10872	12623	+	ATG	TAA	583	5.17	66.47	Hypothetical protein				0	–	–
24	12756	13196	+	ATG	TGA	146	5.59	16.57	Hypothetical protein				0	–	–
25	13171	13545	+	ATG	TGA	124	9.66	14.48	unnamed protein product	Pseudomonas phage OBP (61/118, 52%)	YP_004957952.1	1.00E-36	0	–	–
26	13535	13939	+	ATG	TAA	134	4.41	14.79	unnamed protein product	Pseudomonas phage OBP (72/128, 56%)	YP_004957953.1	6.00E-43	0	–	–
27	14082	15032	+	ATG	TAA	316	4.95	35.36	unnamed protein product	Pseudomonas phage OBP (202/317, 64%)	YP_004957954.1	8.00E-145	0	–	Tubulin/FtsZ, GTPase domain superfamily
28	15366	15842	+	ATG	TGA	158	4.56	18.12	unnamed protein product	Pseudomonas phage OBP (89/159, 56%)	YP_004957956.1	1.00E-52	0	–	–
29	15811	16380	+	ATG	TAA	189	4.73	22.19	unnamed protein product	Pseudomonas phage OBP (54/136, 40%)	YP_004957957.1	2.00E-21	0	–	–
30	16389	16691	+	ATG	TGA	100	4.93	12.11	unnamed protein product	Pseudomonas phage OBP (63/93, 68%)	YP_004957958.1	1.00E-38	0	–	–
31	16672	16989	+	ATG	TAG	105	5.08	11.79	unnamed protein product	Pseudomonas phage OBP (42/96, 44%)	YP_004957959.1	1.00E-24	0	–	–
32	16999	17613	+	ATG	TGA	204	4.14	23.34	unnamed protein product	Pseudomonas phage OBP (64/214, 30%)	YP_004957960.1	7.00E-19	0	–	–
33	17610	18041	+	ATG	TAA	143	4.85	16.56	Putative virion structural protein	Pseudomonas phage OBP (93/139, 67%)	YP_004957961.1	9.00E-63	0	–	–
34	18129	20207	+	ATG	TAA	692	6.85	79.71	T4-like DNA polymerase	Pseudomonas phage OBP (549/692, 79%)	YP_004957962.1	0.00E+00	0	–	SUPERFAMILY SSF56672 (DNA/RNA polymerases)
35	20276	21367	–	ATG	TAA	363	8.86	42.38	Putative virion structural protein	Pseudomonas phage OBP (207/360, 58%)	YP_004957963.1	1.00E-146	0	–	–
36	21625	23541	+	ATG	TAA	638	7.1	70.09	unnamed protein product	Pseudomonas phage OBP (440/613, 72%)	YP_004957964.1	0.00E+00	0	–	–
37	23638	25227	+	ATG	TAA	529	5.96	61.1	RNA polymerase beta subunit	Pseudomonas phage OBP (445/527, 84%)	YP_004957965.1	0.00E+00	0	–	RNA polymerase, alpha subunit (IPR000722
38	25227	25895	+	ATG	TAA	222	4.85	25.43	unnamed protein product	Pseudomonas phage OBP (176/220, 80%)	YP_004957966.1	5.00E-130	0	Yes	–
39	25971	26396	+	ATG	TAA	141	5.92	15.91	unnamed protein product	Pseudomonas phage OBP (70/106, 66%)	YP_004957967.1	4.00E-40	0	–	–
40	26400	26864	+	ATG	TAA	154	5.31	17.34	RNA ligase	Pseudomonas phage OBP (96/150, 64%)	YP_004957968.1	4.00E-64	0	–	–
41	27010	27177	+	ATG	TAA	55	9.99	5.97	unnamed protein product	Pseudomonas phage OBP (31/55, 56%)	YP_004957969.1	5.00E-11	1	–	–
42	27281	27688	+	ATG	TGA	135	5.4	15.33	putative virion structural protein	Pseudomonas phage OBP (100/134, 75%)	YP_004957970.1	8.00E-70	0	–	–
43	27690	28493	+	ATG	TAA	267	4.5	30.65	unnamed protein product	Pseudomonas phage OBP (176/267, 66%)	YP_004957971.1	2.00E-123	0	–	–
44	28520	29656	+	ATG	TGA	378	5.49	43.9	putative SbcD	Pseudomonas phage OBP (292/373, 78%)	YP_004957972.1	0.00E+00	0	–	Metallo-dependent phosphatase-like (IPR029052)
45	29649	30443	+	ATG	TAA	264	9.41	29.56	unnamed protein product	Pseudomonas phage OBP (187/264, 71%)	YP_004957973.1	6.00E-134	0	–	–

46	30458	31225	+	ATG	TAA	255	8.5	28.61	unnamed protein product	Pseudomonas phage OBP (156/201, 78%)	YP_004957974.1	1.00E-109	0	–	–
47	31406	33046	+	ATG	TAA	546	4.9	62.34	unnamed protein product	Pseudomonas phage OBP (425/543, 78%)	YP_004957976.1	0.00E+00	0	–	–
48	33046	34590	+	ATG	TAA	514	4.36	56.38	unnamed protein product	Pseudomonas phage OBP (372/513, 73%)	YP_004957977.1	0.00E+00	0	–	–
49	34613	36085	+	ATG	TAA	490	5.39	53.83	unnamed protein product	Pseudomonas phage OBP (353/505, 70%)	YP_004957978.1	0.00E+00	0	–	–
50	36158	36493	–	ATG	TAA	111	5.61	12.97	unnamed protein product	Pseudomonas phage OBP (80/111, 72%)	YP_004957979.1	1.00E-56	0	–	–
51	36573	38681	+	ATG	TAA	701	8.45	80.33	RNA polymerase beta subunit	Pseudomonas phage OBP (559/694, 81%)	YP_004957980.1	0.00E+00	0	Yes	–
52	38650	40671	+	ATG	TAA	673	6.06	76.62	RNA polymerase beta subunit	Pseudomonas phage OBP (551/673, 82%)	YP_004957981.1	0.00E+00	0	–	–
53	40709	41008	+	ATG	TAA	99	4.9	11.35					0	–	–
54	41008	41310	+	ATG	TAA	100	8.05	11.13	unnamed protein product	Pseudomonas phage OBP (51/93, 55%)	YP_004957984.1	2.00E-29	0	–	–
55	41358	41588	+	ATG	TAA	76	8.98	8.59	unnamed protein product	Pseudomonas phage OBP (43/76, 57%)	YP_004957986.1	6.00E-23	0	–	–
56	41662	41931	+	ATG	TAA	89	5.61	10.18	unnamed protein product	Pseudomonas phage OBP (33/87, 38%)	YP_004957989.1	4.00E-12	0	–	–
57	42089	43633	+	ATG	TAA	514	9.12	59.59	putative helicase	Pseudomonas phage OBP (421/514, 82%)	YP_004957991.1	0.00E+00	0	–	Helicase superfamily 1/2, ATP-binding domain
58	43670	44095	+	ATG	TAA	141	4.4	15.85	unnamed protein product	Pseudomonas phage OBP (53/119, 45%)	YP_004957992.1	7.00E-28	0	–	–
59	44175	44450	+	ATG	TAA	91	4.74	10.73	unnamed protein product	Pseudomonas phage OBP (33/93, 35%)	YP_004957993.1	2.00E-11	0	–	–
60	44470	45030	+	ATG	TAA	186	4.23	20.81	unnamed protein product	Pseudomonas phage OBP (99/181, 55%)	YP_004957994.1	3.00E-53	0	–	–
61	45098	45553	+	ATG	TAA	151	7.74	17.36					0	–	–
62	45697	46146	+	ATG	TAA	149	7.05	16.76					0	–	–
63	46348	47412	+	ATG	TAA	354	6.61	40.41	HNH endonuclease	Erwinia phage vB_EamM_Joad (86/284, 30%)	ASU03793.1	1.00E-20	0	–	HNH nuclease (IPR003615)
64	47455	48762	–	ATG	TAG	435	4.87	25.22	putative virion-associated RNA polymerase beta subunit	Pseudomonas phage OBP (337/436, 77%)	YP_004957997.1	0.00E+00	0	–	–
65	48755	49636	–	ATG	TAA	293	4.19	33.12	putative virion structural protein	Pseudomonas phage OBP (178/295, 60%)	YP_004957998.1	5.00E-124	0	–	–
66	49674	50429	+	ATG	TAA	251	5.93	27.66	unnamed protein product	Pseudomonas phage OBP (101/254, 40%)	YP_004957999.1	3.00E-55	0	–	–
67	50462	52075	+	ATG	TAA	537	5.01	61.68	unnamed protein product	Pseudomonas phage OBP (313/537, 58%)	YP_004958000.1	0.00E+00	0	–	–
68	52186	52518	+	ATG	TAA	110	6.41	12.34	unnamed protein product	Pseudomonas phage OBP (47/108, 44%)	YP_004958001.1	4.00E-25	0	–	–
69	52524	53066	+	ATG	TAA	180	9.24	20.97	unnamed protein product	Pseudomonas phage OBP (91/183, 50%)	YP_004958002.1	1.00E-52	0	–	–
70	53170	53685	+	ATG	TAA	171	8.24	19.53					0	–	–
71	53688	53966	+	ATG	TAA	92	5.84	9.99	unnamed protein product	Pseudomonas phage OBP (64/82, 78%)	YP_004958005.1	3.00E-40	0	–	–
72	54000	54539	+	ATG	TAA	179	5.33	19.67					0	–	–
73	54601	56340	–	ATG	TAA	579	6.15	66.62	T4-like DNA polymerase	Pseudomonas phage OBP (526/582, 90%)	YP_004958006.1	0.00E+00	0	–	Ribonuclease H-like superfamily (IPR012337
74	56429	57772	+	ATG	TAA	447	5.69	49.74	putative virion structural protein	Pseudomonas phage OBP (202/450, 45%)	YP_004958007.1	6.00E-132	0	–	–
75	57782	58216	+	ATG	TAA	144	4.79	16.63	putative virion structural protein	Pseudomonas phage OBP (84/144, 58%)	YP_004958008.1	2.00E-51	0	–	–
76	58231	59811	+	ATG	TAA	526	5.98	58.94	putative virion structural protein	Pseudomonas phage OBP (256/525, 49%)	YP_004958009.1	5.00E-177	0	–	–
77	59813	60466	+	ATG	TAA	217	5.13	25.18	unnamed protein product	Pseudomonas phage OBP (108/217, 50%)	YP_004958010.1	1.00E-71	0	–	–
78	60514	63501	–	ATG	TAA	995	5.16	112.86	putative virion structural protein	Pseudomonas phage OBP (695/994, 70%)	YP_004958011.1	0.00E+00	0	–	–
79	63510	64595	–	ATG	TAA	361	5.03	41.19	putative virion structural protein	Pseudomonas phage OBP (289/359, 81%)	YP_004958012.1	0.00E+00	0	–	–
80	64622	65710	+	ATG	TAA	362	6.05	39.42	putative minor virion structural protein	Pseudomonas phage OBP (270/363, 74%)	YP_004958013.1	0.00E+00	0	–	–
81	65713	66597	+	ATG	TAA	294	5.45	33.54	putative virion structural protein	Pseudomonas phage OBP (174/294, 59%)	YP_004958014.1	2.00E-124	0	–	–
82	66581	67111	+	ATG	TAG	176	6.43	19.98	putative virion structural protein	Pseudomonas phage OBP (113/176, 64%)	YP_004958015.1	8.00E-82	0	–	–
83	67101	68384	+	ATG	TAA	427	4.79	48.37	putative virion structural protein	Pseudomonas phage OBP (254/420, 60%)	YP_004958016.1	0.00E+00	0	–	–
84	68374	69336	+	ATG	TAA	320	6.27	35.85	hypothetical protein PPEV_gp062	Pseudomonas phage EL (81/316, 26%)	YP_418095.1	1.00E-21	0	–	–
85	69374	70597	+	ATG	TAA	407	5.47	42.66	putative virion structural protein	Pseudomonas phage OBP (128/416, 31%)	YP_004958019.1	8.00E-47	0	–	–
86	70706	71767	+	ATG	TGA	353	9.08	39.94	putative virion structural protein	Pseudomonas phage OBP (168/358, 47%)	YP_004958020.1	6.00E-101	0	–	–
87	71764	72276	+	ATG	TGA	170	4.99	19.09	unnamed protein product	Pseudomonas phage OBP (67/171, 39%)	YP_004958021.1	1.00E-34	0	–	–
88	72273	72881	+	ATG	TGA	202	4.78	22.96	unnamed protein product	Pseudomonas phage OBP (66/201, 33%)	YP_004958022.1	1.00E-14	0	–	–
89	72850	73689	+	ATG	TAA	279	5.57	30.44	unnamed protein product	Pseudomonas phage OBP (130/281, 46%)	YP_004958023.1	7.00E-73	0	–	–
90	73691	74398	+	ATG	TAA	235	5.57	26.69	unnamed protein product	Pseudomonas phage OBP (88/233, 38%)	YP_004958024.1	9.00E-44	0	–	–
91	74661	76247	+	ATG	TGA	528	5.08	62.35	putative virion structural protein	Pseudomonas phage OBP (350/526, 67%)	YP_004958025.1	0.00E+00	0	–	–
92	76216	77517	+	ATG	TAA	433	5.16	50.32	putative virion structural protein	Pseudomonas phage OBP (294/433, 68%)	YP_004958026.1	0.00E+00	0	–	–
93	77528	78166	+	ATG	TAA	212	5.57	24.26	putative virion structural protein	Pseudomonas phage OBP (130/211, 62%)	YP_004958027.1	2.00E-90	0	–	–
94	78166	79509	+	ATG	TAA	447	5.34	50.47	putative virion structural protein	Pseudomonas phage OBP (293/445, 66%)	YP_004958028.1	0.00E+00	0	–	–
95	79558	81096	–	ATG	TAA	512	5.26	58.07	putative DNAB-like replicative helicase	Pseudomonas phage OBP (437/512(85%)	YP_004958029.1	0.00E+00	0	–	–
96	81134	81685	+	ATG	TAA	183	5.55	21.18	unnamed protein product	Pseudomonas phage OBP (126/183, 69%)	YP_004958030.1	2.00E-89	0	–	–
97	81802	83979	+	ATG	TAA	725	5.52	78.93	putative major capsid protein	Pseudomonas phage OBP (563/720, 78%)	YP_004958031.1	0.00E+00	0	–	–
98	84397	85230	+	ATG	TAA	277	6.21	31.32	unnamed protein product	Pseudomonas phage OBP (166/276, 60%)	YP_004958032.1	3.00E-106	0	–	–
99	85243	86937	+	ATG	TAA	564	8.7	65.36	unnamed protein product	Pseudomonas phage OBP (497/563, 88%)	YP_004958033.1	0.00E+00	0	–	–
100	86941	89352	+	ATG	TAA	803	5.23	92.04	unnamed protein product	Pseudomonas phage OBP (442/825, 54%)	YP_004958034.1	0.00E+00	0	–	–

101	89568	89996	+	ATG	TAA	142	9.41	16.59	unnamed protein product	Pseudomonas phage OBP (32/69, 46%)	YP_004958035.1	5.00E-13	0	–	–
102	90008	90190	+	ATG	TAG	60	3.93	7.14	unnamed protein product	Pseudomonas phage OBP (42/60, 70%)	YP_004958036.1	7.00E-19	0	–	–
103	90144	90533	+	ATG	TAA	129	5.36	14.95					0	–	–
104	90623	91285	+	ATG	TAA	220	4.64	23.51					0	–	–
105	91346	94234	–	ATG	TAA	962	4.96	108.25	putative portal protein	Pseudomonas phage OBP (656/943, 70%)	YP_004958039.1	0.00E+00	0	–	–
106	94263	96383	+	ATG	TAG	706	5.6	82.05	putative virion structural protein	Pseudomonas phage OBP (505/702, 72%)	YP_004958040.1	0.00E+00	0	–	–
107	96376	96714	+	ATG	TAA	112	6.51	13.17	unnamed protein product	Pseudomonas phage OBP (73/108, 68%)	YP_004958041.1	6.00E-42	0	–	–
108	96721	97653	+	ATG	TAA	310	5.01	34.28	putative major virion structural protein	Pseudomonas phage OBP (230/303, 76%)	YP_004958042.1	1.00E-168	0	–	–
109	97673	98290	+	ATG	TAA	205	5.74	23.32	unnamed protein product	Pseudomonas phage OBP (100/193, 52%)	YP_004958043.1	3.00E-63	0	–	–
110	98295	98888	+	ATG	TAA	197	5.65	22.25	putative RuvC	Pseudomonas phage OBP (145/188, 77%)	YP_004958044.1	3.00E-105	0	–	Ribonuclease H superfamily
111	98848	99498	+	ATG	TAA	216	9.1	23.34	unnamed protein product	Pseudomonas phage OBP (141/218, 65%)	YP_004958045.1	4.00E-95	4	–	–
112	99551	99799	–	ATG	TAG	82	4.55	9.13	lambda Rz1-like protein	Pseudomonas phage OBP (43/67, 64%)	YP_004958046.1	1.00E-22	0	–	–
113	99846	100295	–	ATG	TGA	149	7.84	16.61	lambda Rz1-like protein	Pseudomonas phage OBP (75/146, 51%)	YP_004958047.1	7.00E-40	1	–	–
114	100292	101521	–	ATG	TGA	409	7.2	47.37	putative virion structural protein	Pseudomonas phage OBP (273/408, 67%)	YP_004958048.1	0.00E+00	0	–	–
115	101551	102315	+	ATG	TAA	254	5.15	28.29	putative virion structural protein	Pseudomonas phage OBP (169/254, 67%)	YP_004958049.1	7.00E-124	0	–	–
116	102379	106098	+	ATG	TAA	1239	5.84	140.82	putative virion structural protein	Pseudomonas phage OBP (561/1266, 44%)	YP_004958050.1	0.00E+00	0	–	–
117	106159	109923	+	ATG	TAA	1254	5.44	141.6	putative virion structural protein	Pseudomonas phage OBP (625/1278, 49%)	YP_004958051.1	0.00E+00	0	–	–
118	109967	115864	+	ATG	TAA	1965	5.54	221.36	putative virion structural protein	Pseudomonas phage OBP (807/1745, 46%)	YP_004958052.1	0.00E+00	0	–	Protein of unknown function DUF285
119	115915	116910	–	ATG	TAA	331	5.09	35.87	putative phage baseplate assembly protein	Pseudomonas phage OBP (221/330, 67%)	YP_004958053.1	6.00E-166	0	–	–
120	116932	119412	+	ATG	TAA	826	6.86	94.44	putative SbcC ATPase	Pseudomonas phage OBP (531/826, 64%)	YP_004958054.1	0.00E+00	0	–	P-loop containing nucleoside triphosphate hydrolase
121	119451	120008	+	ATG	TAA	185	4.94	20.89	unnamed protein product	Pseudomonas phage OBP (76/187, 41%)	YP_004958055.1	2.00E-38	0	–	–
122	120226	121035	+	ATG	TAA	269	9.92	30.38	putative lytic murein transglycosylase	Pseudomonas phage OBP (135/258, 52%)	YP_004958056.1	4.00E-87	1	Yes	Lysozyme-like domain superfamily
123	121097	121360	+	ATG	TAA	87	5.41	9.43	unnamed protein product	Pseudomonas phage OBP (42/87, 48%)	YP_004958057.1	2.00E-21	0	–	–
124	121427	122149	+	ATG	TAG	240	8.98	27.53					0	–	–
125	122157	124130	+	ATG	TAA	657	5.27	73.98	putative NAD-dependent DNA ligase	Pseudomonas phage OBP (347/651, 53%)	YP_004958060.1	0.00E+00	0	–	NAD-dependent DNA ligase
126	124190	124513	+	ATG	TAA	107	5.61	12.07					0	–	–
127	124513	125088	+	ATG	TAA	191	5.88	21.65					0	–	–
128	125341	125712	+	ATG	TAA	123	9.25	14.17	putative ribonuclease H	Pseudomonas phage OBP (90/108, 83%)	YP_004958065.1	2.00E-58	0	–	Ribonuclease H-like superfamily
129	125740	127539	+	ATG	TAA	599	4.44	66.71	unnamed protein product	Pseudomonas phage OBP (116/202, 57%)	YP_004958068.1	5.00E-59	0	–	–
130	127676	128989	+	ATG	TAA	437	5.47	49.72					0	–	–
131	129132	129878	+	ATG	TAA	248	6.31	28.25	putative homing nuclease	Pseudomonas phage OBP (137/245, 56%)	YP_004958072.1	3.00E-91	0	–	GIY-YIG endonuclease
132	129884	130171	+	ATG	TAA	95	9.93	10.15					0	–	Lambda repressor-like, DNA-binding domain superfamily
133	130164	130529	+	ATG	TAA	121	9.75	13.71					0	–	Lambda repressor-like, DNA-binding domain superfamily
134	130547	131530	+	ATG	TGA	327	7.03	37.72	unnamed protein product	Pseudomonas phage OBP (68/195, 35%)	YP_004958076.1	1.00E-21	0	–	–
135	131530	131889	+	ATG	TAA	119	9.1	13.46	unnamed protein product	Pseudomonas phage OBP (39/115, 34%)	YP_004958077.1	2.00E-15	0	–	–
136	131889	132581	+	ATG	TAA	230	6.92	25.56					0	–	–
137	132615	133376	+	ATG	TAA	253	9.11	28.71	unnamed protein product	Pseudomonas phage OBP (56/208, 27%)	YP_004958081.1	3.00E-06	0	–	–
138	133376	133663	+	ATG	TAA	95	5.69	10.78	hypothetical protein	Acinetobacter phage ABPH49 (23/62, 37%)	AXN57887.1	9.00E-05	0	Yes	–
139	133894	134316	+	ATG	TAA	140	9.74	15.85					0	–	–
140	134316	134543	+	ATG	TAA	75	9.16	8.79					2	–	–
141	134566	134706	+	ATG	TAA	46	9.82	5.09	hypothetical protein EHM17_17180	Verrucomicrobiaceae bacterium (32/43, 74%)	RPJ30187.1	5.00E-13	0	–	–
142	134766	135125	+	ATG	TAA	119	7.86	13.68					0	–	–
143	135112	135453	+	ATG	TGA	113	5.37	13.1	unnamed protein product	Pseudomonas phage OBP (40/109, 37%)	YP_004958087.1	2.00E-13	0	–	–
144	135450	135734	+	ATG	TAA	94	9.7	10.47					1	Yes	–
145	135745	136203	+	ATG	TAA	152	7.77	17.27	unnamed protein product	Pseudomonas phage OBP (39/155, 25%)	YP_004958089.1	2.00E-04	0	–	–
146	136282	136500	+	ATG	TAA	72	9.1	8.32					0	–	–
147	136493	137872	+	ATG	TAA	459	5.75	50.72	unnamed protein product	Pseudomonas phage OBP (160/505, 32%)	YP_004958090.1	3.00E-37	0	–	–
148	137884	138576	+	ATG	TAA	230	5.01	26.16	unnamed protein product	Pseudomonas phage OBP (151/225, 67%)	YP_004958091.1	5.00E-108	0	–	–
149	138704	139018	+	ATG	TAA	104	6.07	12.36					0	–	–
150	139048	139533	+	ATG	TGA	161	9.27	18.61					0	–	–
151	139533	139745	+	ATG	TAA	70	9.7	8.41					0	–	–
152	139756	139995	+	ATG	TAA	79	10.61	9.16					0	–	–

153	140045	140674	+	ATG	TAA	209	5.73	24.07	unnamed protein product	Pseudomonas phage OBP (122/205, 60%)	YP_004958100.1	1.00E-85	0	–	–
154	140674	141009	+	ATG	TAA	111	5.91	12.85	unnamed protein product	Pseudomonas phage OBP (85/110, 77%)	YP_004958101.1	3.00E-57	0	–	–
155	141040	141387	+	ATG	TAA	115	5.01	12.8					0	–	–
156	141387	141713	+	ATG	TAA	108	10.39	12.34					0	–	–
157	141727	143034	+	ATG	TAA	435	5.94	48.6	unnamed protein product	Pseudomonas phage OBP (37/110, 34%)	YP_004958106.1	3.00E-05	0	–	–
158	143065	143412	+	ATG	TAA	115	6.71	12.24	unnamed protein product	Pseudomonas phage OBP (48/116, 41%)	YP_004958109.1	1.00E-26	1	Yes	–
159	143473	143661	–	ATG	TAA	62	8.01	6.94					0	–	–
160	143755	145923	–	ATG	TAA	722	4.55	78.47	unnamed protein product	Pseudomonas phage OBP (60/150, 40%)	YP_004958112.1	6.00E-22	0	–	–
161	145938	149954	–	ATG	TAA	1338	5.24	144.28	unnamed protein product	Pseudomonas phage OBP (148/492, 30%)	YP_004958113.1	6.00E-42	0	–	–
162	150579	160706	–	ATG	TAA	3375	4.77	372.47	unnamed protein product	Pseudomonas phage OBP (1062/2918, 36%)	YP_004958114.1	0.00E+00	0	–	–
163	160820	163636	–	ATG	TAA	938	6.1	105.58	unnamed protein product	Pseudomonas phage OBP (203/973, 21%)	YP_004958115.1	2.00E-09	0	–	–
164	163715	164380	–	ATG	TAA	221	5.4	24.83	unnamed protein product	Pseudomonas phage OBP (72/204, 35%)	YP_004958116.1	1.00E-25	0	–	–
165	164495	165292	–	ATG	TAA	265	9.09	29.76	putative HNH homing endonuclease	Erwinia phage vB_EamM_Special G (113/244, 46%)	ANJ64902.1	1.00E-66	0	–	HNH nuclease
166	165344	166477	–	ATG	TAA	377	4.94	43.24	putative NrdB	Pseudomonas phage OBP (311/377, 82%)	YP_004958118.1	0.00E+00	0	–	Ribonucleotide reductase small subunit family
167	166781	167635	+	ATG	TAA	284	5.26	31.73	unnamed protein product	Pseudomonas phage OBP (49/126, 39%)	YP_004958122.1	3.00E-19	0	Yes	–
168	167790	168416	+	ATG	TAA	208	9.07	22.83					1	–	–
169	168432	168827	+	ATG	TAA	131	9.81	14.73					1	–	–
170	169056	169331	+	ATG	TGA	91	5.71	10.36	MULTISPECIES: hypothetical protein	Burkholderia	WP_023842450.1	5.00E-08	1	–	–
171	169352	169639	+	ATG	TAA	95	8.68	10.11					1	Yes	–
172	169723	170124	+	ATG	TAA	133	6.3	15.42	unnamed protein product	Pseudomonas phage OBP (71/133, 53%)	YP_004958125.1	9.00E-48	0	–	–
173	170126	170410	+	ATG	TAA	94	5.44	11.11	hypothetical protein	Aeromonas phage SD04 (35/74, 47%)	AWY07228.1	9.00E-07	1	–	–
174	170619	170831	+	ATG	TAA	70	5.25	8.24					0	–	–
175	170842	171006	+	ATG	TAA	54	10.69	6.04					0	–	–
176	170996	171766	+	ATG	TAA	256	5.83	29.77	unnamed protein product	Pseudomonas phage OBP (75/237, 32%)	YP_004958130.1	5.00E-29	0	–	–
177	171777	172196	+	ATG	TAA	139	4.85	16.28	unnamed protein product	Pseudomonas phage OBP (55/146, 38%)	YP_004958131.1	2.00E-21	0	–	–
178	172196	172654	+	ATG	TAA	152	5.37	17.78	unnamed protein product	Pseudomonas phage OBP (63/153, 41%)	YP_004958132.1	2.00E-36	0	–	–
179	172725	173330	–	ATG	TAA	201	5.15	22.06	unnamed protein product	Pseudomonas phage OBP (111/200, 56%)	YP_004958133.1	3.00E-74	0	–	–
180	173450	173905	+	ATG	TAA	151	8.49	17.8	unnamed protein product	Pseudomonas phage OBP (61/125, 48%)	YP_004958134.1	4.00E-05	0	–	–
181	173881	174285	+	ATG	TAA	134	4.86	15.86					0	–	–
182	174297	174824	+	ATG	TGA	175	7.17	20.22	putative serine/threonine protein phosphatase	Pseudomonas phage OBP (123/167, 74%)	YP_004958136.1	4.00E-93	0	–	Metallo-dependent phosphatase-like
183	174799	175026	+	ATG	TAA	75	11.43	8.83					2	–	–
184	175231	176421	+	ATG	TAA	396	6.35	46.49	putative thymidylate synthase	Pseudomonas phage OBP (247/396, 62%)	YP_004958141.1	0.00E+00	0	–	Thymidylate synthase/dCMP hydroxymethylase superfamily
185	176501	176725	+	ATG	TAA	74	6.8	8.63					2	–	–
186	176736	177635	+	ATG	TAA	299	8.64	32.95					0	–	–
187	177776	178126	+	ATG	TAA	116	8.01	12.92	unnamed protein product	Pseudomonas phage OBP (53/115, 46%)	YP_004958144.1	1.00E-24	0	–	–
188	178213	178860	+	ATG	TAA	215	5.41	25.39	unnamed protein product	Pseudomonas phage OBP (77/201, 38%)	YP_004958146.1	9.00E-36	0	–	HAD-like superfamily. Phosphoglycolate phosphatase-like, domain
189	178863	179486	+	ATG	TAA	207	8.54	23.37	putative dihydrofolate reductase	Pseudomonas phage OBP (92/201, 46%)	YP_004958148.1	1.00E-47	0	–	Dihydrofolate reductase
190	179503	179757	+	ATG	TAA	84	10.02	9.31	unnamed protein product	Pseudomonas phage OBP (24/59, 41%)	YP_004958149.1	6.00E-10	0	–	–
191	179851	179940	+	ATG	TAG	29	10.02	3.45					0	–	–
192	179971	180618	+	ATG	TGA	215	5.56	24.4					0	–	Clp protease proteolytic subunit /Translocation-enhancing protein T
															ATP-dependent Clp protease proteolytic subunit
193	180645	181037	+	ATG	TAA	130	9.53	15.07	hypothetical protein pf16_08	Pseudomonas phage pf16 (29/59, 49%)	AND74931.1	1.00E-04	2	–	–
194	181086	182645	+	ATG	TAA	519	5	57.41	putative chaperonin GroEL	Pseudomonas phage OBP (339/541, 63%)	YP_004958153.1	0.00E+00	0	–	Chaperonin Cpn60/TCP-1 family. Chaperone tailless complex polype
195	183006	183287	+	ATG	TAA	93	8.66	10.96	unnamed protein product	Pseudomonas phage OBP (50/88, 57%)	YP_004958155.1	2.00E-31	1	–	–
196	183304	183888	+	ATG	TGA	194	8.19	21.83	thymidine kinase	Psychromonas sp. psych-6C06 (120/198, 61%)	WP_101107339.1	2.00E-86	0	–	Thymidine kinase
197	183979	184326	+	ATG	TAA	115	5.68	11.65	unnamed protein product	Pseudomonas phage OBP (38/82, 46%)	YP_004958157.1	1.00E-08	0	–	–
198	184764	185486	+	ATG	TAA	240	5.28	27.65	unnamed protein product	Pseudomonas phage OBP (91/233, 39%)	YP_004958159.1	2.00E-28	0	–	–
199	185617	186108	+	ATG	TAA	163	4.55	18.07	unnamed protein product	Pseudomonas phage OBP (59/136, 43%)	YP_004958161.1	3.00E-08	0	–	–
200	186108	186599	+	ATG	TAA	163	9.77	18.74					0	–	–
201	186599	187303	+	ATG	TAA	234	5.9	26.97	unnamed protein product	Pseudomonas phage OBP (66/138, 48%)	YP_004958163.1	8.00E-22	0	–	–
202	187331	187906	+	ATG	TAA	191	8.89	22.65	unnamed protein product	Pseudomonas phage OBP (58/196, 30%)	YP_004958164.1	6.00E-17	0	–	–
203	187964	188602	–	ATG	TAA	212	4.68	23.92	putative virion structural protein	Pseudomonas phage OBP (146/216, 68%)	YP_004958165.1	1.00E-117	0	–	–

204	188599	189930	–	ATG	TGA	443	5.14	50.58	putative virion structural protein	Pseudomonas phage OBP (339/447, 76%)	YP_004958166.1	0.00E+00	0	–	–
205	189985	190389	+	ATG	TAG	134	7.92	16	unnamed protein product	Pseudomonas phage OBP (89/136, 65%)	YP_004958167.1	1.00E-65	1	–	–
206	190450	191871	+	ATG	TAA	473	8.22	54.23	unnamed protein product	Pseudomonas phage OBP (364/456, 80%)	YP_004958168.1	0.00E+00	0	–	–
207	191928	193016	–	ATG	TAA	362	5.52	39.72	putative virion structural protein	Pseudomonas phage OBP (124/379, 33%)	YP_004958169.1	7.00E-19	0	–	–
208	193020	195119	–	ATG	TAA	699	6.37	81.32	putative virion-associated SNF2-domain containing helicase	Pseudomonas phage OBP (513/706, 73%)	YP_004958170.1	0.00E+00	0	–	P-loop containing nucleoside triphosphate hydrolase
209	195137	196240	+	ATG	TAA	367	4.59	42.1	unnamed protein product	Pseudomonas phage OBP (182/348, 52%)	YP_004958171.1	2.00E-106	0	–	–
210	196289	197068	–	ATG	TAA	259	5.72	28.89	putative virion structural protein	Pseudomonas phage OBP (152/260, 58%)	YP_004958172.1	7.00E-118	0	–	–
211	197136	198605	+	ATG	TAA	489	6.01	55.81	putative UvsX	Pseudomonas phage OBP (425/491, 87%)	YP_004958173.1	0.00E+00	0	–	–
212	198607	198969	+	ATG	TGA	121	4.56	13.71	unnamed protein product	Pseudomonas phage OBP (78/12, 65%)	YP_004958174.1	2.00E-62	0	–	–
213	198969	199646	+	ATG	TAA	225	5.56	24.74	putative virion structural protein	Pseudomonas phage OBP (184/225, 82%)	YP_004958175.1	1.00E-170	0	–	–
214	199646	200122	+	ATG	TAA	158	4.69	18.22	unnamed protein product	Pseudomonas phage OBP (79/141, 56%)	YP_004958176.1	5.00E-52	0	–	–
215	200129	200857	+	ATG	TAA	242	5.53	28.16	unnamed protein product	Pseudomonas phage OBP (171/246, 70%)	YP_004958177.1	1.00E-151	0	–	–
216	200906	201508	–	ATG	TAA	200	4.28	22.64	unnamed protein product	Pseudomonas phage OBP (137/204, 67%)	YP_004958178.1	2.00E-112	0	–	–
217	201618	202169	+	ATG	TGA	183	5.1	21.25	unnamed protein product	Pseudomonas phage OBP (114/189, 60%)	YP_004958179.1	7.00E-86	0	–	–
218	202162	202572	+	ATG	TAA	136	9.52	15.66	unnamed protein product	Pseudomonas phage OBP (99/136, 73%)	YP_004958180.1	8.00E-70	0	–	–
219	202642	203757	+	ATG	TAA	371	4.51	40.46	putative virion structural protein	Pseudomonas phage OBP (255/369, 69%)	YP_004958181.1	0.00E+00	0	–	–
220	203816	206020	–	ATG	TAA	734	5.42	79.74	putative virion structural protein	Pseudomonas phage OBP (578/740, 78%)	YP_004958182.1	0.00E+00	3	–	–
221	206060	213247	–	ATG	TAA	2395	8.14	260.37	putative structural lysozyme	Pseudomonas phage OBP (1376/2449, 56%)	YP_004958183.1	0.00E+00	0	–	Lysozyme-like domain superfamily
222	213278	214894	+	ATG	TAA	538	5.41	60.58	putative virion-associated RNA polymerase beta subunit	Pseudomonas phage OBP (454/519, 87%)	YP_004958184.1	0.00E+00	0	–	–
223	214894	219234	+	ATG	TAA	1446	5.49	163.48	putative virion-associated RNA polymerase beta subunit	Pseudomonas phage OBP (1057/1451, 73%)	YP_004958185.1	0.00E+00	0	–	–
224	219237	220103	+	ATG	TAA	228	8.75	32.28	hypothetical protein PPEV_gp188	Pseudomonas phage EL (114/287, 40%)	YP_418221.1	3.00E-52	0	–	–
225	220131	220298	+	ATG	TGA	55	4.57	6.16	putative virion structural protein	Pseudomonas phage OBP (41/50, 82%)	YP_004958187.1	4.00E-20	0	–	–
226	220295	221812	+	ATG	TAA	505	4.66	56.69	putative virion structural protein	Pseudomonas phage OBP (300/504, 60%)	YP_004958188.1	0.00E+00	0	–	–
227	221812	222549	+	ATG	TAA	245	5.19	27.26	putative virion structural protein	Pseudomonas phage OBP (196/245, 80%)	YP_004958189.1	9.00E-144	0	–	–
228	222551	223399	+	ATG	TAA	282	5.57	31.91	putative virion structural protein	Pseudomonas phage OBP (205/283, 72%)	YP_004958190.1	3.00E-153	0	–	–
229	223503	224102	+	ATG	TAA	199	5.09	22.56	putative virion structural protein	Pseudomonas phage OBP (94/159, 59%)	YP_004958191.1	4.00E-63	0	–	–
230	224210	224596	+	ATG	TAA	128	4.79	14.25	unnamed protein product	Pseudomonas phage OBP (47/115, 41%)	YP_004958198.1	4.00E-15	0	–	–
231	224603	225136	+	ATG	TGA	177	5.79	20.26					0	–	–
232	225387	225803	+	ATG	TAA	138	5.04	15.73	unnamed protein product	Pseudomonas phage OBP (44/130, 34%)	YP_004958198.1	8.00E-15	0	–	–
233	225913	226365	+	ATG	TAA	150	5.17	16.61					0	–	–
234	226405	226842	+	ATG	TAA	145	4.98	16.48	unnamed protein product	Pseudomonas phage OBP (96/147, 65%)	YP_004958202.1	2.00E-59	0	–	–
235	226907	227308	+	ATG	TAA	133	10.21	15.64					0	–	–
236	227259	227669	+	ATG	TAA	136	8.99	16.29	unnamed protein product	Pseudomonas phage OBP (68/128, 53%)	YP_004958206.1	1.00E-44	0	–	–
237	227660	228118	+	ATG	TAA	152	8.61	17.72	unnamed protein product	Pseudomonas phage OBP (38/98, 39%)	YP_004958207.1	2.00E-13	0	–	–
238	228118	228720	+	ATG	TGA	200	6.98	23.44	unnamed protein product	Pseudomonas phage OBP (140/200, 70%)	YP_004958208.1	1.00E-104	0	–	–
239	228795	229334	+	ATG	TAA	179	5.7	19.74	unnamed protein product	Pseudomonas phage OBP (76/186, 41%)	YP_004958209.1	3.00E-36	1	Yes	–
240	229431	230072	+	ATG	TAA	213	5.49	23.06	unnamed protein product	Pseudomonas phage OBP (79/215, 37%)	YP_004958210.1	9.00E-34	1	Yes	–
241	230112	230786	+	ATG	TAA	224	5.29	24.24	unnamed protein product	Pseudomonas phage OBP (102/228, 45%)	YP_004958213.1	5.00E-64	0	Yes	–
242	230931	231635	+	ATG	TAA	234	5.4	27.17	putative metal-dependent phosphohydrolase	Pseudomonas phage OBP (171/233, 73%)	YP_004958214.1	3.00E-121	0	–	–
243	232568	233470	–	ATG	TAA	300	5.98	33.49	putative tail tube	Pseudomonas phage OBP (271/300, 90%)	YP_004957908.1	0.00E+00	0	–	–
244	233482	235677	–	ATG	TAA	731	4.92	80.78	putative major tailsheath	Pseudomonas phage OBP (561/733, 77%)	YP_004957909.1	0.00E+00	0	–	–
245	235785	236660	+	ATG	TAA	291	8.4	32.27	putative virion structural protein	Pseudomonas phage OBP (233/291, 80%)	YP_004957910.1	1.00E-177	0	–	–
246	236688	239501	+	ATG	TGA	937	5.22	107.84	putative virion structural protein	Pseudomonas phage OBP (667/938, 71%)	YP_004957911.1	0.00E+00	0	–	–
247	239494	241236	+	ATG	TAA	580	5.44	65.13	putative virion structural protein	Pseudomonas phage OBP (443/578, 77%)	YP_004957912.1	0.00E+00	0	–	–
248	241285	244695	–	ATG	TAA	1136	8.38	130.91	putative terminase large subunit	Pseudomonas phage OBP (483/599, 81%)	YP_004957913.1	0.00E+00	0	–	Homing endonuclease. Hint domain superfamily
249	244825	245451	–	ATG	TAA	208	5.85	23.43	putative thymidylate kinase	Pseudomonas phage OBP (78/199, 39%)	YP_004957915.1	9.00E-39	0	–	Thymidylate kinase
250	245451	246917	–	ATG	TAA	488	6.2	56.88	putative virion structural protein	Pseudomonas phage OBP (379/484, 78%)	YP_004957916.1	0.00E+00	0	–	–
251	246917	247594	–	ATG	TAA	225	7.72	24.98	putative virion structural protein	Pseudomonas phage OBP (189/225, 84%)	YP_004957917.1	1.00E-138	0	–	–
252	248377	248889	+	ATG	TGA	170	9.1	18.92					0	–	–
253	248953	249192	–	ATG	TAA	79	4.99	8.61	putative glutaredoxin	Pseudomonas phage OBP (46/64, 72%)	YP_004957919.1	9.00E-24	0	–	Thioredoxin-like superfamily
254	249816	250403	–	ATG	TAA	195	9.47	21.69	hypothetical protein SHAb15599_00126	Acinetobacter phage SH-Ab 15599 (58/150, 39%)	AXF41482.1	2.00E-31	0	–	–
255	251200	251439	–	ATG	TAG	79	5.73	9.1	hypothetical protein SopranoGao_55	Klebsiella phage SopranoGao (32/73, 44%)	ASV45078.1	2.00E-10	0	–	Protein of unknown function DUF2829
256	252260	252874	+	ATG	TAA	204	7.72	23.63	putative virion structural protein	Pseudomonas phage OBP (62/197, 31%)	YP_004957923.1	1.00E-28	0	–	–
257	252889	253245	+	ATG	TAG	118	7.76	13.99	unnamed protein product	Pseudomonas phage OBP (40/62,65%)	YP_004957924.1	3.00E-10	1	–	–

Supplementary table 4. MEME predicted promotors in *Klebsiella aerogenes* phage N1M2 shared with Pseudomonas phage OBP.

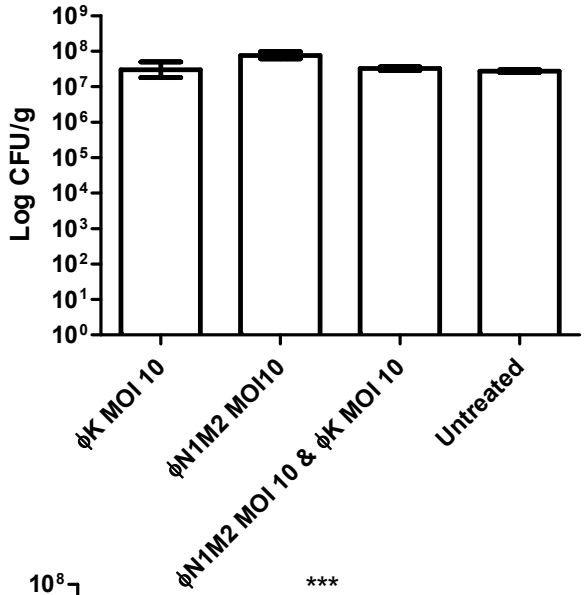
Promoter	Start	End	Sequence	E value
			TBYAWWWWWTTTCARRYAKATATTATYWAAGTGWA	3.5e-093
pORF 9	4117	4152	CTCATCAATAATCAAATCGAGATCTTGAAAATCTT	8.9e-5
pORF 14	6381	6417	CTTATTTGTTTTGAAATAGATATTATTAGATTGAA	1.6e-10
pORF 24	12701	12735	GTTATAAAATTTCAAACAAATATTATCCTCATGTA	7.4e-12
pORF 41	26928	26963	TGTATTTCATTTCAGACCTATATTACTTAAGTGAT	1.0e-12
pORF 47	31335	31370	TCTAAAAATGTTTAGGCAGATATTATTAGAGTGTA	3.1e-14
pORF 50	36515 complement	36550	TACATATAGATAATATCTGTCTGATTATCTTTGGC	5.5e-13
pORF 51	36515	36550	GCCAAAGATAATCAGACAGATATTATCTATATGTA	5.5e-13
pORF 56	41620	41655	GTAATAAATTTTTAAGTAGATATTATTTATAGGAG	3.0e-10
pORF 70	53113	53149	TCTATTGATTTTCAAGCAGATATTATTAACATGTA	4.6e-15
pORF 85	69322	69357	AACAATTTAATTAAATCGTTTTAACTCTTTCTATA	4.0e-6
pORF 104	90554	90589	TGAAAGAAATTTCAGGTCTATATCATCTAAGTGTA	6.8e-12
pORF 122	120146	120181	GTCACATATTTTCAGGCAGATATTATCATTATGTA	2.8e-13
pORF 131	129052	129087	TGCATATAAGTTTAAGCAGATATTATCACAGTGTA	2.7e-14
pORF 149	138627	138662	TGTATTTATTTTTGGACTTATATTATCTAAGTGTA	1.7e-12
pORF 167	166693	166728	TCCAAATCTTTTCAGGTAGATATTATTAAAGTGAA	8.6e-16
pORF 168	167710	167745	TGCATTTCATTTTAAACCGATATTATTAACGTGAG	2.3e-13
pORF 179	173383 complement	173418	TACATAATGATAATATCTGTCTATTTTTATTTGAA	2.1e-11
pORF 180	173383	173418	TTCAAATAAAAATAGACAGATATTATCATTATGTA	2.1e-11
pORF 187	177684	177720	TGTATTTCTTTTCAGACATATATAATTAAAGTGTA	8.5e-15
pORF 192	179906	179941	GGTAAAATTTTTCAAACAGATATTATCTAAGTAGT	4.0e-12
pORF 197	183895	183930	TGTATCTTATTTCAGACATATATCATTTAAGTGAA	2.5e-13
pORF 199	185535	185570	TCTATTTTAAAACAGATAGATATTATCATAGTGTA	6.1e-13
pORF 216	201530 complement	201565	TCTACTTTGATAATATTTGTTTGAAATATAATAGT	3.4e-11
pORF 217	201530	201565	ACTATTATATTTCAAACAAATATTATCAAAGTAGA	3.4e-11
pORF 219	202605	202640	CTCTTTTAAATAATTTATTTTTTTTTAGGATTAAT	3.7e-5
pORF 230	224133	224168	GCTATTTAATTTCAAGCAGATATTATTAACGTGTT	2.0e-15
pORF 233	225820	225854	GCCAAAATGCTTCAGGTATATATTATTACCGTGAA	1.3e-11
pORF 239	228726	228761	TACATTCATAGTCAGGTCTATATTATCTAAGTGTA	1.0e-11
pORF 240	229336	229371	TGAATACAAATTCAGAAATATATAATCTAATTGAT	3.2e-10
pORF 242	230861	230895	TTTAAAATTATTCAGACGTATATTATTAGCATGTA	4.0e-12
pORF 256	252179	252214	TCCATTATTAAATAAATAGATATTATTAGATTGAA	6.0e-11

Supplementary table 5. Terminators of Phage N1M2 detected using ARNold and confirmed using Mfold Quikfold. Loops and stems are shown in red and blue respectively.

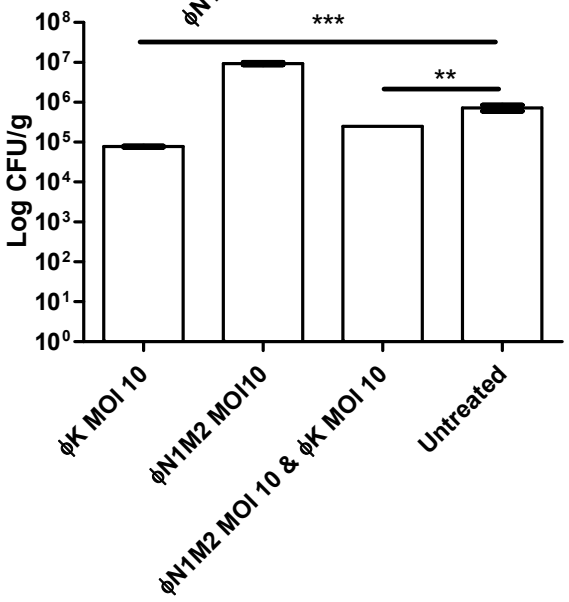
Terminator	Start	End	Sequence	ΔG kcal/mol
tORF5	2731	2770	GGGGTGGTGGGTAAAACCGTCACCCCgTTTTGATACCAC	-18.60
tORF13	6347	6384	CTATCACCTCCAAAAGGGGTGATAGTTTTTAGTTCTT	-14.10
tORF14	7153	7189	TTGGTCTGGTGGTTATCAGATCAATTTACTTTCTGG	-10.10
tORF18	9195	9229	TGGGGAGGTCAACCCTCCCTAaTTTTAATAGATC	-13.12
tORF27	15041	15072	GCCCTGGCTAGTCCAGGGCTTTTTATATTGC	-13.60
tORF32	17679	17710	GCGCTCAATACAAGAGCGCTTTGATTACCGG	-10.50
tORF	23568 complement	23612	CAGAGAGGAGGGCTTTCACCCACCCTCTCTGTTAATTTAGTTAA	-17.20
tORF36	23582	23625	CAGAGAGGGTGGGTGAAAGCCCTCCTCTCTGTTGTTTTTATAGG	-23.20
tORF38	25903	25941	GCAATGGGGTGGAAACACCCTGTTGCTTTTATCAGGTT	-21.40
tORF	27171 complement	27213	CAGGATCTCCCGGTTAGGGGAGATCCTGgTTATGACTTATTT	-22.00
tORF41	27185	27226	CAGGATCTCCCCTAACCGGGAGATCCTGTTATTCTATACTT	-20.60
tORF50	36095 complement	36135	TACCACCTACTATAGAGTAGGTGGTAgTTATTTACATAAA	-17.40
tORF60	45037	45076	CTTACCCTTCCTAGTGGAGGGTAAGggTTATATGTTTAT	-14.80
tORF	46140 complement	46179	CTACTCACCCTTAATTGGGTGAGTAGgTATGTGTTATTT	-18.90
tORF64	47405 complement	47446	CCTCTACTCCCGTGATGGGAGTAGAGGTTATGTTTACTTA	-20.60
tORF63	47419	47457	CTCTACTCCCATCACGGGAGTAGAGgTCTGTTATGCTA	-18.40
tORF67	52161	52193	GCCGTCTTAGAAAGGCGGCTTTTCATGATCA	-11.10
tORF73	54535 complement	54578	CCCTACCCACTCCGAAAGGAGGGGTAGGGgTTATGTTTGTTAT	-27.00
tORF85	70629	70667	GAGTAGGGGCCTTCGGGCTCCTACTCTTATATGTTTGC	-24.10
tORF87	72330	72366	GCTTGGCTGCTAAACAAGCTAAGCTATCTGTTAATG	-12.00
tORF90	74598	74637	CTCTACTCCTTTTATCGGGAGTAGGGgTTTGATGTTATG	-15.10
tORF	94012	94044	CAGGTCATTATTGGCCTGaaTGTTATTACTGA	-10.20
tORF108	97725	97671	GACCACGGTTTCGGCCGTGGTCagTTCTATTTATGA	-17.10
tORF109	98249	98288	TGTATGGGGGTTATAAAACCCATGCAcTTTTTGTAGGAT	-11.30
tORF111	99509	99554	CCCTACCCCTCCAATTAAGGATGGGGTAGGGgTTATGCTTACTAT	-23.50
tORF	109916 complement	109945	GGGGAGTTTCCTCCCCaTTCTATTAACCT	-11.50
tORF117	109929	109957	GGGGAGGAAACTCCCCTTTCTTATTTAG	-15.40
tORF119	115859 complement	115903	CCCCTACCCCTCCGAAAGGAAGGGTAGGGAgTTTATGGTTTATA	-24.50
tORF118	115875	115915	CCTACCCTTCCTTTCGGAGGGGTAGGggTTATGTTTACTT	-21.00
tORF120	119412	119438	CCCCGGAAACGGGGTTTTATTCATTA	-12.70
tORF125	124134	127172	TAGGAGAGGGGAAACTCTCTCCTAcaTTTTTGTTTATA	-19.10
tORF127	125095	125135	CCTAGAGAGGGAATTCCCTCTCTAGGacTTTATATTGTTT	-22.00
tORF	134705 complement	134741	GAACCTACCTTAATCGGTAGGTTCTCTATTATGGTT	-14.90
tORF	14139 complement	141438	TGGGGGTTGGTTTTACCCAACCCTTGaTTCTTTCTAATA	-13.40
tORF161	150437	150467	GGGGACCGGAGTCCCCCCTTTTACATCAAC	-10.70
tORF163	160772 complement	160813	TGAGAGTCTGGGGTAACCTGGGCTCTCATTATTATGGTATG	-19.50
tORF164	163667 complement	163705	AGTCCCTGGGTTCGTCCAGGGACTcTTTATCAGATTCT	-18.90
tORF165	164443 complement	164479	GTAGGGTGGCCGAAGCCACCCTACTCTATGTTCGGA	-19.80
tORF171	169645	169686	AGAGAGGCTTCCGAAAGGAGGTCTCTCTcTTTTTGTACCTGA	-22.70
tORF179	172660 complement	172704	CCCTACCTAGCCCGCAAAGGCTAGGTAGGGaTTATGCATGTATT	-24.40

tORF184	176430	176468	GCTAGGGGCTCTTAGGAGCCTTTAGCTTTTCAAGGTAT	-17.50
tORF	182639 complement	182678	CGAGCTGAGGTGTTACCCTCAGCTCGaTCTTAGTTATGC	-19.00
tORF	184319 complement	184355	CAGGCTACCCGAAGGTAGCCTGagTTTACTTATTCG	-16.30
tORF197	184344	184638	AGGCTACCTTCGGGTAGCCTgTTATTCGGTTCTT	-17.10
tORF203	187904 complement	187944	CCTACCCTTCCTTACGGAGGGGTAGGggTATTTTGTCGTT	-21.00
tORF206	191881	191920	CTTACCCTCCGCAAGGAAGGGTAAGggTTATATGTTTGT	-18.00
tORF210	196236 complement	196271	CTTCCTTGCGGAGGGTAAGGGAGTATGTTTGTTAT	-11.30
tORF220	203754 complement	203798	CCCTTACCCACTCCGTGAGGAGGGGTAAGGGTTATGCTTGTTTA	-27.20
tORF221	206022 complement	206057	TGAGGGGACTTCGGTCCCCTTAcTTTTAGAGGTAA	-18.10
tORF	214038 complement	214070	CTGTGACGTTATCGTCGCAGTTTTCTACTGAT	-11.60
tORF223	219297 complement	219328	AGGCCGTCAATGCGGCCTTTATATAGGGAAA	-11.00
tORF	222630 complement	222667	GTTGGGGATATTAAAACCCCCAGCgTTTAATTCATAC	-11.80
tORF228	223407	223443	GTAGGGGTCCTTCGGGATCTCTACTTATTTTGTTAT	-18.20
tORF243	232501 complement	232545	CCCCTACCCACTCCGAAAGGAGGGGTAGGGGTTTTATGTTGTTT	-29.60
tORF	247875 complement	247919	CCCTACTCCCTCCGAAAGGAAGGGGTAGGGgTTTATGTTTGTTT	-24.60
tORF	251572 complement	251608	GCCAGAGAAATGAATCTCTGGCacTTTTAATGTGAT	-14.40

A.



B.



Supplementary Figure 1. Effect of N1M2 and Phage K on mixed biofilms formed by N1 *K. aerogenes* and DPC 5247 *S.aureus*. No glucose 48 hr biofilm, 48 hr phage. N1 and DPC 5247 were quantified in the biofilm by dilution and spread plating on UTI ChromoSelect agar. The experiment was carried out once but counts were carried out in duplicate. A. Count of N1 in mixed biofilm formed by N1 and DPC 5247. B. Count of DPC 5247 in mixed biofilm formed by N1 and DPC 5247.

General discussion

Phages have myriad uses with some used more classically, such as in studies of co-evolution, and some due to the changing needs of the world. The changing attitudes of consumers towards traditional chemicals, a move towards more natural options, and the increase in resistance to well characterised antimicrobial compounds has reinvigorated the study of phages.

We showed that a commercially available phage reduced *Listeria monocytogenes* growth in coleslaw over a 10 day period at both high multiplicity of infection (MOI) and low MOI. The low MOI alone was as effective as a combination of the phage and a bacteriocin but the combination may be helpful in limiting the emergence of phage resistant mutants. Although listeriosis is a serious illness many consumers are unaware of it or are uninformed about its mode of transmission. In a survey 78.9% of pregnant women reported awareness of *L. monocytogenes* but conversely 71.1% reported consuming high-risk food, suggesting a gap between recognition of a term and full understanding of the associated risks (1). This was despite the fact that pregnant women are ten times more likely to contract listeriosis than the general population and 14% of all listeriosis cases occur in pregnant women. In a separate study 80% of non-pregnancy associated listeriosis patients in Germany had never heard of the disease prior to their illness (2). Also only 6% of the immunocompromised control population knew about the foodborne nature of the disease. *L. monocytogenes* can grow at high salt concentrations, a wide pH range, and at refrigeration temperatures. *L. monocytogenes* are found in a range of environments including a wide diversity of foods, soil, water, vegetation, sewage, and animal faeces (3). The ability of *L. monocytogenes* to form biofilms is also important for its

persistence in the environment, including processing plants and contaminated food, and its subsequent spread (4). Phage and bacteriocins are being introduced as clean label antimicrobials in food processing. As mentioned in Chapter 1b the properties of foods can affect the activity of phages so the combination could be tested in foods of different pHs, textures, and at different storage temperatures. Phage concentration is vital to the success of phage application in food so a range of concentrations were investigated. Partial synergy between the two biological agents was achieved in the checkerboard assays but not in the food model but this could be investigated further by varying the concentrations used. As also discussed phage choice is important so an alternative phage, such as ListShield, could be applied. Similarly, variants of nisin with increased activity could be applied to the system (5).

No differences were found between the viromes of men and women or between any of the phases of the menstrual cycle. No differences were visible in the bacteriome as analysed by 16S rRNA data. Differences were visible between the viromes of women using hormonal contraception and women not using hormonal contraception. We believe this to be the first study of its kind. No published literature is available on the effect of hormonal contraception on the gut virome or bacteriome assessed by metagenomics sequencing. Inter-individual variation in the viromes of people and viral dark matter can mask differences between groups or individuals. Clustering was used to overcome this issue. Clustered virome data showed more pronounced differences than unclustered data. The gut microbiota can regulate hormones and hormones in turn can affect the gastrointestinal tract (6). The menstrual cycle, and its associated changes in hormone levels, can be associated with gastrointestinal symptoms (7). Changes in the menstrual cycle have been found in the year prior to irritable bowel disease diagnosis (8). The selection of healthy controls is vital to

establishing differences between groups in metagenomics studies. A number of factors must be taken into account when recruiting healthy controls. Further research should be carried out to determine if there are differences between users and non-users of hormonal contraceptives. This study was a pilot study carried out with a limited number of subjects and must be repeated with greater numbers. Also metagenomic sequencing of the bacteriome could establish if there were any differences between any of the groups and if links were visible between the bacterial and phage components of the microbiome that were not visible using 16S data. This could help to establish if the virome is more discriminatory than the bacteriome or if they reflect each other (9). Comparisons of the effect of different oral contraceptive types such as estrogen and progesterone combined and progesterone only should be carried out. Postmenopausal women receiving and not receiving hormone replacement therapy could also be investigated and a comparison of premenopausal and postmenopausal women.

When a healthy human faecal sample was screened for phages against 163 bacterial strains isolated from the same sample no phages were isolated. One phage was isolated from the faecal sample against a well characterised strain after screening an additional panel of 77 bacterial strains available in the laboratory. However, 155 phages were isolated from six environmental samples against the laboratory strains and the strains isolated from a human faecal sample. This highlights the importance of screening new and unrelated environments and of intelligent strain selection for phage screening (10). Of the 86 phages sequenced the majority were previously published, which draws attention to the difficulty in isolating new and novel phages (11). Single nucleotide polymorphisms (SNPs) analysis was carried out on 70 strains of a previously described phage compared to the original genome and to one another. Although these were strains of the same phage genetic differences were apparent.

SNPs introduced synonymous and non-synonymous changes and introduced stop codons. Some genes were found to only contain synonymous changes or low levels of SNPs which could indicate the need for conservation of these genes. It would be of interest to screen a sample for phages using a method including and excluding enrichment to establish how much more effective enrichment can be. However, it must be taken in to account that enrichment can bias isolation. A systematic review of literature could be carried out to try and establish the probability of isolating phages based on the number of sources screened, the number of strains used, the number of potential phages found, if enrichment was included, and the number of novel phages found in a multitude of studies. It would be interesting to assess the host range of the APCEc01 phage strains including against the original host and pinpoint which, if any, SNPs affected this. Transcriptomics or ribosome profiling could be used to try and establish roles for the uncharacterised, hypothetical proteins or phage proteins present in APCEc01 (12). Some genes which had a large number of SNPs had no annotated function so little could be learned from these areas. SNPs are commonly investigated during co-evolution experiments where phages and bacteria are cultured together for a number of generations. SNPs have previously been investigated by designing PCR primers to regions of DNA, amplifying the region, and then Sanger sequencing the region (13). Another option is using host range as an indicator by assessing infectivity towards the original host and hosts from different time points in the evolution (14). Bacteria can also be assessed in the same way. Whole genome sequencing as carried out in this study is also an option.

A novel lytic jumbo phage, N1M2, targeting a *Klebsiella aerogenes* strain was isolated from maize silage. N1M2 could also infect a *Klebsiella pneumoniae* strain. The *K. aerogenes* strain was isolated from a healthy human faecal sample. *K.*

aerogenes is an ESKAPE pathogen and often found as a component of biofilms on medical devices. N1M2 at MOIs of 10 and 100 applied for 48 hours was not effective against biofilms formed over 48 hours. However, N1M2 applied at an MOI of 10 or 100 for 72 hours significantly reduced a biofilm formed over 48 hours and similarly N1M2 applied at an MOI of 10 or 100 for 48 hours significantly reduced a biofilm formed over 24 hours. Phage K is a well-documented phage with activity against *Staphylococcus aureus* (15). Similar to *K. aerogenes S. aureus* is of great importance for antibiotic resistance and is found as biofilms on medical devices. In a mixed community biofilm of *K. aerogenes* and *S. aureus* grown for 48 hours N1M2 alone and Phage K alone at MOIs of 10 applied for 48 hours did not significantly reduce the biofilm. However, N1M2 and Phage K at MOIs of 10 in combination significantly reduced the biofilm. Antibiotic resistance is naturally occurring but rates have increased due to the over use and misuse of antibiotics in humans and animals (16). Antibiotics have been used as growth enhancers in livestock, as prophylactics for infections, and to treat already established infections. New treatments are required for treatment of bacterial infections due to this rapid and unrelenting rise. Phage therapy is a technique that has largely been unexplored in Western medicine but is now gaining interest. Further characterisation of N1M2 could be carried out including its stability at a range of temperatures and pHs, host range analysis using more bacterial strains, restriction enzyme analysis, and a one-step growth curve. Host range should be assessed using efficiency of plating to avoid false positives in the form of lysis from without. The antibiotic sensitivity of strain N1 could be assessed. N1 was sequenced using the MinION and was described as a draft genome and should be sequenced further using a different platform. *K. aerogenes* resistance to carbapenem antibiotics is commonly caused by the overexpression of efflux pumps, the modification of

porins, or the downregulation of porin synthesis (17). Biofilm assays should be carried out to compare the efficacy of antibiotics and N1M2 against *K. aerogenes* biofilms. Combinations of antibiotics and phage could also be investigated. This could include the eradication of pre-formed biofilms and inhibiting the formation of biofilms. Biofilm assays could also be performed with clinical isolates from biofilms or infected wounds.

This thesis addressed a varied range of applications of phages. Phage was applied to food in combination with another natural food additive, a pilot study highlighted the need for further investigation in to the effect of hormonal contraception on the human gut virome and bacteriome, a novel jumbo phage was isolated and tested in a model system, and the difficulties associated with screening and single nucleotide polymorphisms in the genomes of multiple versions of a phage were investigated.

References

1. **Xu W, Cater M, Gaitan A, Drewery M, Gravois R, Lammi-Keefe CJ.** 2017. Awareness of *Listeria* and high-risk food consumption behavior among pregnant women in Louisiana. Food Control **76:**62-65.

2. **Maia RL, Teixeira P, Mateus TL.** 2019. Risk communication strategies (on listeriosis) for high-risk groups. Trends in Food Science & Technology **84:**68-70.

3. **Todd ECD, Notermans S.** 2011. Surveillance of listeriosis and its causative pathogen, *Listeria monocytogenes*. Food Control **22:**1484-1490.

4. **Smith MK, Draper LA, Hazelhoff P-J, Cotter PD, Ross RP, Hill C.** 2016. A Bioengineered Nisin Derivative, M21A, in Combination with Food Grade Additives Eradicates Biofilms of *Listeria monocytogenes*. Frontiers in Microbiology **7:**1939.

5. **Field D, Begley M, O'Connor PM, Daly KM, Hugenholtz F, Cotter PD, Hill C, Ross RP.** 2012. Bioengineered Nisin A Derivatives with Enhanced Activity against Both Gram Positive and Gram Negative Pathogens. PLoS ONE **7:**e46884.

6. **Chen KL, Madak-Erdogan Z.** 2016. Estrogen and Microbiota Crosstalk: Should We Pay Attention? Trends in Endocrinology & Metabolism **27:**752-755.

7. **Bharadwaj S, Barber MD, Graff LA, Shen B.** 2015. Symptomatology of irritable bowel syndrome and inflammatory bowel disease during the menstrual cycle. Gastroenterology Report **3:**185-193.

8. **Saha S, Zhao Y, Shah SA, Esposti SD, Lidofsky S, Salih S, Bright R, Law M, Moniz H, Flowers N, Merrick M, Sands BE.** 2014. Menstrual Cycle Changes in Women with Inflammatory Bowel Disease: A Study from the Ocean State Crohn's and Colitis Area Registry. Inflammatory bowel diseases **20:**534-540.

9. **Broecker F, Russo G, Klumpp J, Moelling K.** 2017. Stable core virome despite variable microbiome after fecal transfer. Gut Microbes **8**.

10. **McLaughlin MR, Balaa MF, Sims J, King R.** 2006. Isolation of *Salmonella* bacteriophages from swine effluent lagoons. J Environ Qual **35:**522-528.

11. **Callaway TR, Edrington TS, Brabban A, Kutter E, Karriker L, Stahl C, Wagstrom E, Anderson RC, Genovese K, McReynolds J, Harvey R, Nisbet DJ.** 2010. Occurrence of *Salmonella*-specific bacteriophages in swine feces collected from commercial farms. Foodborne Pathog Dis **7:**851-856.

12. **Danis-Wlodarczyk K, Blasdel BG, Jang HB, Vandenheuvel D, Noben J-P, Drulis-Kawa Z, Lavigne R.** 2018. Genomic, transcriptomic, and structural analysis of *Pseudomonas* virus PA5oct highlights the molecular complexity among Jumbo phages. bioRxiv doi:10.1101/406421**:**406421.

13. **Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD.** 2013. Rapid evolution of the human gut virome. Proceedings of the National Academy of Sciences **110:**12450-12455.

14. **Scanlan PD, Hall AR, Lopez-Pascua LD, Buckling A.** 2011. Genetic basis of infectivity evolution in a bacteriophage. Mol Ecol **20:**981-989.

15. **Alves DR, Gaudion A, Bean JE, Perez Esteban P, Arnot TC, Harper DR, Kot W, Hansen LH, Enright MC, Jenkins ATA.** 2014. Combined Use of Bacteriophage K and a Novel Bacteriophage To Reduce *Staphylococcus aureus* Biofilm Formation. Applied and Environmental Microbiology **80:**6694-6703.

16. **Gordillo Altamirano FL, Barr JJ.** 2019. Phage Therapy in the Postantibiotic Era. Clinical Microbiology Reviews **32:**e00066-00018.

17. **Lavigne J-P, Sotto A, Nicolas-Chanoine M-H, Bouziges N, Pagès J-M, Davin-Regli A.** 2013. An adaptive response of *Enterobacter aerogenes* to imipenem: regulation of porin balance in clinical isolates. International Journal of Antimicrobial Agents **41:**130-136.

Acknowledgements

Thank you to Prof. Colin Hill and Prof. Paul Ross for the opportunity to do a PhD.

Thank you to everyone in the gut phageomics and bacteriocins labs. As Albus Dumbledore said “Happiness can be found, even in the darkest of times, if one only remembers to turn on the light” and oftentimes the people around the lab were that light. I really believe it’s the people that make somewhere a nice place to work not the work. Thanks for all of the birthday cakes, biscuit days, treats for absolutely no reason, and for loving cake as much as I do. I would like to especially thank Lorraine and Stephen. Lorraine, you deserve a medal for all you do for us.

To Maddie, Fiona, and Shónagh you are wonderful friends and I am so lucky you are in my life. Norah, Kieran, and Sinead you get a mention here too. And now to Caitriona. In all the time there was never a moment that you doubted that I would finish and I want to thank you for that. There was many a time that I didn’t think the end would ever come but you never waivered in your belief and friendship. To all of you thanks for caring about stupid things and very, very important things.

Thank you to my family Eva, Paul, David, Laurence, Sarah, Gill, Conor, and Isabel. I appreciate all of the times that you talked about and didn’t talk about my PhD. Isabel and Conor thank you for always cheering me up, every problem seems so insignificant when you are around.

A very special thank you to my parents, Maeve and Willy, who have put up with me and my moods during this time. I could not have done it without your love and support. I don’t know how to thank you for all you have done.